

A New Benchmark Based on Recent Advances in Multispectral Pansharpening

Revisiting pansharpening with classical and emerging pansharpening methods

GEMINE VIVONE, MAURO DALLA MURA, ANDREA GARZELLI, ROCCO RESTAINO, GIUSEPPE SCARPA, MAGNUS O. ULFARSSON, LUCIANO ALPARONE, AND JOCELYN CHANUSSOT

Pansharpening refers to the fusion of a multispectral (MS) image and panchromatic (PAN) data aimed at generating an outcome with the same spatial resolution of the PAN data and the spectral resolution of the MS image. In the last 30 years, several approaches to deal with this issue have been proposed. However, the reproducibility of these methods is often limited, making the comparison with the state of the art hard to achieve. Thus, to fill this gap, we propose a new benchmark consisting of recent advances in MS pansharpening. In particular, optimized classical approaches [multiresolution analysis (MRA) and component substitution (CS)] are compared with methods belonging to the third generation of pansharpening, represented by variational optimization-based (VO) and machine learning (ML) techniques. The benchmark is tested on different scenarios (from urban to rural areas) acquired by different commercial sensors [i.e., *IKONOS (IK)*, *GeoEye-1 (GE-1)*, and *WorldView-3 (WV-3)*]. Both quantitative and qualitative assessments and the computational burden are analyzed in this article, and all of the implementations have been collected in a MATLAB toolbox that is made available to the community.

OVERVIEW OF PANSHARPENING

The main objective of satellite remote sensing is to provide accurate reproductions of the Earth's surface. This can be obtained by improving hardware to resolve many more details in both space and frequency domains. However, it turns out to be a hard task because of the stringent constraint of the signal-to-noise ratio of satellite products. A way around the problem consists of combining multiple images with complementary features to get high-quality products through signal processing [1].

Pansharpening, which stands for *panchromatic sharpening*, is a powerful and well-known image fusion methodology involving the fusion of an MS and a PAN image to provide a final outcome with the same spatial resolution as the PAN image and the same spectral resolution as the MS image. Pansharpening has proven to be useful for both visual interpretation (see, e.g., its use in commercial software such as Google Earth and Bing Maps) and as a preliminary step for higher-level processing (see, e.g., the use of pansharpened images for crop mapping [2], land cover monitoring [3], and anomaly detection [4]). Some pansharpening techniques have also been extended to slightly different issues, i.e., the fusion of PAN and hyperspectral (HS) images [5], [6] and the combination of MS and HS data [7]–[10].

Digital Object Identifier 10.1109/MGRS.2020.3019315
Date of current version: 29 October 2020

BACKGROUND AND RELATED WORKS

The central role played by pansharpening methods within remotely sensed image fusion techniques is testified to by a comprehensive book on the subject [11] as well as review articles as [12]. Actually, specifically devoted surveys have populate the literature in recent years [13]–[17], corroborating the growing interest in this topic.

Several taxonomies have been applied to pansharpening algorithms. For each case, relationships among the different classes can be found. In this article, we adopt a broad classification that is pointed to practical use. In particular, we defined four main classes characterized by well-distinguished properties that are crucial in the selection of the most appropriate algorithm for the specific application.

The first two classes (CS and MRA) make up the group of classical approaches, which was deeply analyzed in a previous study [14]. However, as will be widely illustrated in the following, significant improvements have been obtained in recent years, motivating the introduction of a large number of novel methods in this article. Indeed, since the recent developments preserve the key characteristic of high computational efficiency, classical algorithms are still of fundamental importance.

Instead, the two remaining classes (VO and ML) represent the main emerging lines of research in the last few years [18], [19]. Their unequivocal objective is to achieve the uncompromised quality of the fused product, which often induces a higher level of complexity and a consequent increase of computational effort.

We start with a brief description of the four classes and go on to describe the specific methods selected for implementation and assessment. Finally, we review the literature related to the debated problem of the quantitative assessment of fused products.

COMPONENT SUBSTITUTION

The CS approaches are also referred to as *spectral* methods. They are based on the projection of the original MS image in a transformed domain. The objective is to ease the replacement, either partial or total, of the spatial information with that of the PAN image. Thanks to their simple implementation, many pioneering pansharpening algorithms belong to the CS class. For example, those exploiting intensity-hue-saturation (IHS) [20], [21] and principal component analysis (PCA) [22], [23] date back to the early 1990s. However, the appealing visual features of the final products and the robustness against spatial misalignments have favored the development of several other algorithms belonging to this class. Other methods, for example, the Gram-Schmidt (GS) expansion [24], the matting model [25], nonlinear IHS [26], and nonlinear PCA [27], have been exploited for deriving more specific image representations with an easily identifiable spatial component.

Considering the different image transforms, a plethora of combination algorithms have been proposed for

introducing PAN spatial information into the original MS data. The main contribution of the recent literature has been the improvement of the injection rules to the extent that the new methods can be considered second-generation techniques. The focus of these studies has centered on accurately modeling the relationship between the pixel values of the PAN image and those of the MS channels. An iterative fitting algorithm was proposed in [28], but the most used methods rely on statistical modeling. This methodology was implemented, in different forms, by the partial replacement adaptive CS (PRACS) [29] and the adaptive GS (GSA) [30]. A more comprehensive study can be found in [31].

Another fruitful line of research has originated in the local application of CS algorithms. This principle has been used by completing the fusion task pixel by pixel, using a rectangular sliding window for calculating the involved quantities [30], or by performing the pansharpening process by preliminarily identified image segments [32].

Noticeably, the detail extraction and injection phases are not always considered in a separate way. For example, the band-dependent spatial detail (BDS) method is based on the contemporary estimation of the weights for constructing a low-pass version of the PAN image from the MS channels and of the injection coefficients [33]. Also, in this case, significant improvements have been obtained by partitioning the image through a segmentation algorithm [34] and deeply analyzing the statistical relationship among the available images [35].

MULTIREOLUTION ANALYSIS

The class of MRA methods is composed of approaches that utilize a multiscale decomposition. In most instances, the MRA is directly applied to the signals in the original spatial domain, and, thus, another name is the class of *spatial* methods. Also, in this case, it is possible to trace these methods back to fusion approaches based on general-purpose decompositions, performed by means of a simple low-pass filter, or by more complex iterative approaches, such as Laplacian pyramids [36], wavelets [37]–[39], curvelets [40], and contourlets [41]. These techniques were applied to the pansharpening problem more than ten years ago [21], [42]–[46].

However, due to the appreciable spectral quality of the fused images achievable through MRA approaches, much effort has been devoted in subsequent years to the design of more specific techniques. A crucial step that can be used to define the beginning of the second-generation era is represented by the introduction of information regarding the acquisition sensor into the decomposition schemes [47], [48]. Adaptive techniques dealing with the accuracy of knowledge about the sensor model and with the specificity of the considered images were suggested in [49]–[53].

Recent improvements focus on the utilization of different decomposition schemes and the optimization of the

injection coefficients. Indeed, leveraging on the success in other image processing areas, nonlinear approaches have also been applied to pansharpening; examples include least-squares support vector machine [54] and morphological filters (MF) [55]. Furthermore, significantly improved injection procedures have been designed by means of a more thorough analysis of the relationship between the available images [31], [56], [57] and the effects of the atmosphere on the collected signal [56], [58]–[60].

The MRA class does not complete the set of classical methods, since an important group is made up of hybrid approaches that combine the CS and MRA approaches. Although some examples of MRA+CS (MRA followed by CS) methods exist [31], the most important hybrid techniques adopt the CS+MRA (CS followed by MRA) scheme; namely, they consist of applying an MRA decomposition in a transformed domain [56], [61], [62]. Accordingly, they can be recast into the MRA class [48]. The classical additive wavelet luminance proportional (AWLP) [61], recently improved in [56] by considering the characteristics of the acquisition device and the optical radiation transfer process, is the most widespread example. Using independent component analysis in conjunction with curvelets [63] and using PCA in combination with contourlets [46] or guided filters [64] are other noticeable instances.

VARIATIONAL OPTIMIZATION-BASED TECHNIQUES

VO techniques are pansharpening methods that rely on the solution of an optimization problem. They have gained popularity in recent years thanks to advances in convex optimization and inverse problems in imaging, such as super-resolution, blind deconvolution, and image restoration. All techniques in this family rely on a model that, for example, can describe the process of acquisition done by a sensor or the representation of an image. Such models lead to the definition of an objective function that will be treated in general by a variational optimization approach for finding a suitable solution. More specifically, in several VO techniques, the relationship between the observed PAN image, the low-resolution MS (LRMS) images, and the high-resolution MS (HRMS) image is established according to a sensor model. The problem to be solved, then, is to estimate the HRMS images from the PAN and LRMS images. In this framework, the pansharpening problem is ill-conditioned, which means that a direct inversion will cause noise amplification. To mitigate the ill conditioning, some kinds of regularization are necessary. The estimation problem in a variational setting involves a cost function that generally consists of two terms, i.e., the fidelity term that describes the relationship between the HRMS images and the observed images and a regularization term that incorporates some prior beliefs about the HRMS images into the optimization process. The HRMS images are the solution to this optimization problem. The methods described in the following assume that the LRMS images and the PAN image

are properly registered, which is a typical assumption for other approaches as well. We note, however, that pansharpening methods dealing with misregistered data have been proposed [65].

The seminal VO pansharpening article is [66], which proposed the so-called P + XS method based on three assumptions: A1) the LRMS images are assumed to be a low-pass filtered and subsampled version of the desired HRMS images, A2) the PAN image is assumed to be a linear combination of the channels of the HRMS images, and A3) the geometric information of an image is contained in its level sets. To enforce A3, they introduced a regularization term based on the total variation (TV) penalty incorporating geometric details of the PAN image. The A1 and A2 assumptions were incorporated into the fidelity terms. The authors of [67] assumed A1 and A2 but employed a TV regularizer, separately, on each band of the HRMS images. In [68], information about the PAN image was incorporated into the TV regularizer. Furthermore, a low-rank structure was assumed in the spectral domain. The authors of [69] relaxed A2 by imposing it only for high-frequency components and used a roughness that includes both the HRMS images and the PAN image. The authors of [70] and [71] share a regularizer with the P + XS method but propose different fidelity terms. The authors of [72] did not directly assume A2 but indirectly modeled the relationship between the PAN and the HRMS images by considering a spectral low-rank relationship between them, using an l_2 roughness penalty to regularize the solution. In [50], an algorithm is proposed to directly estimate the spatial relationship between the PAN and the MS images. Other notable articles include [73]–[76].

Bayesian methods also fall into the VO category. The authors of [77] assumed a joint Gaussian model for the unknown MS image and the PAN method; [78] extended this idea by incorporating wavelet transformation into the formulation; [79] further extended this model using PCA to enforce a low-rank spectral structure. Other notable Bayesian articles include [76] and [80].

Another class of methods belonging to the VO family relies on sparse representations [81]. The idea is to represent the unknown HRMS image as a sparse linear combination of dictionary elements (here, image patches). The main challenge is the generation of the dictionary, with contributions ranging from using available PAN and LRMS images [82] to using the data at hand [83]–[87] to create the dictionary. In [88], sparse representation was used to reconstruct only the high-resolution details to be injected into LRMS.

Most of the VO methods described previously depend on one or more regularization parameters. These parameters need to be selected by the user. This critical problem is mostly unexplored in the pansharpening literature, and most articles report what settings were used in the experiments. However, some articles provide guidelines, e.g., [72], [86].

MACHINE LEARNING

ML has moved its early steps toward pansharpening in the last decade. Pioneering works in this category include compressive sensing or dictionary-based solutions, such as [81], [83]–[86], [88]. Needless to say, the deep learning (DL) wave has even flooded over into the remote sensing field [19], including pansharpening [89]–[100] and intimately related tasks, such as superresolution [101]–[103] or HS/MS data fusion [104], [105]. The first attempt by Huang et al. in 2015 relies on a modified sparse denoising autoencoder scheme [89]. In 2016, Zhong et al. [90] instead proposed a CS approach based on the GS transform integrated with an off-the-shelf superresolution convolutional neural network (CNN) [106] to upsample the MS component. In parallel, in 2016, Masi et al. [91] designed and trained the first fully convolutional network specifically conceived for pansharpening, the pansharpening neural network (PNN), achieving state-of-the-art results. Motivated by the encouraging results of the PNN, many researchers have moved on to this line of inquiry. In particular, the use of CNNs with residual learning modules [107] will soon become a standard option for pansharpening, as testified to by many subsequent works [92]–[100]. In contrast to the relatively simple architecture of the PNN and its advanced version, advanced PNN (A-PNN) [95], other works explore more complex models in terms of depth [92], [94], [99] and/or topology [97], [98]. Different training strategies have also been investigated by means of different target loss functions, such as the mean squared error (MSE) [91], [92], [96]–[98], the mean absolute error [95], [99] or the *erreur relative globale adimensionnelle de synthèse* (ERGAS) [100], also using adaptive tuning schemes (A-PNN) or resorting to the generative adversarial learning paradigm [99].

A common trait of these ML approaches is the assumption of a training paradigm relying on a resolution downgrade process (e.g., Wald's protocol) to provide labeled data for supervised training. By doing so, due to the scale sensitivity of the pansharpening problem, spatial consistency was soon recognized as a primary issue for ML methods. In fact, the latest ML solutions seek to address such a problem using different training paradigms, in particular, by means of multiobjective strategies. In [108] a joint reduced-/full-scale loss is proposed to enforce spatial fidelity thanks to the introduction of an additional loss term computed at the target scale. Another interesting example is the Pan-GAN model [109], which relies on spectral–spatial multiobjective adversarial training using two dedicated discriminators. It is also worth mentioning other HS/MS fusion approaches, straightforwardly extendable to pansharpening, that include unsupervised training schemes to avoid the resolution downgrade of the data to be used for training [104], [105].

ASSESSMENT

Quality assessment of MS pansharpening has been the object of extensive research activities. Since the target of image

fusion is unavailable, several protocols of quality evaluation have been developed to overcome the lack of a reference image.

A widely adopted protocol was first proposed in [110] and further discussed in [111] and [112]. Such a protocol relies on two properties that the fused data have to satisfy, referred to as *consistency* and *synthesis*, which are reviewed in the “Quality Assessment of Fusion Products” section, together with the indexes that can be adopted for the quantitative assessment of such properties. It should be noted that the consistency property can be checked by comparing the original MS image to a degraded version of the pansharpened image obtained through modulation transfer function (MTF) decimation filters tuned to the MS sensor. This property, which represents a necessary condition to be satisfied by any pansharpening algorithm, has been recently considered sufficient to assess the fusion quality [113]. Conversely, the synthesis property cannot be directly verified, since it would require the target image, which is not available. Therefore, synthesis is checked at degraded spatial scales according to reduced resolution (RR) assessment. The original MS image acts as the reference image to be compared to the fusion product obtained from spatially degraded MS and PAN images. This assessment is accurate, thanks to the presence of a reference MS image, but its main assumption of quality invariance across scales is rarely satisfied. Furthermore, the way the original MS and PAN images are degraded can significantly bias the quality assessment.

As an alternative to RR quality assessment, the problem of measuring the quality of fusion can be approached at the full spatial scale without any spatial degradation, i.e., at full resolution (FR). The spectral and spatial distortions are separately evaluated from the original LRMS bands and the high-resolution PAN image, as originally proposed in [114]. The spectral distortion is calculated for each band as the average absolute difference between the fused band and the corresponding interpolated original band, while the spatial quality is measured by the correlation coefficient between the spatial details of each of the fused MS bands and those of the PAN image.

A widely adopted FR assessment is based on the quality with no reference (QNR) index [115], which combines into a unique overall quality index a spectral distortion measure between the original and pansharpened MS bands and a spatial distortion measure between each MS band and PAN at FR and RR. Another FR protocol, proposed in [116], borrows the spectral consistency property from [110] and the spatial quality from the matching of high-pass spatial details as in [114]. Recently, the hybrid QNR (HQNR) was presented in [117] as the combination of the spectral distortion index proposed in [116] with the spatial distortion index of the original QNR [115]. Moreover, an overall quality metric, exploiting the spectral distortion index as in [115] and a spatial distortion index based on the natural

TABLE 1. A LIST OF THE MAIN SYMBOLS USED IN THIS ARTICLE.

SYMBOL	DESCRIPTION
MS	MS image
\overline{MS}	MS image interpolated at PAN scale
P	PAN image
\widehat{MS}	Pansharpened image
R	Spatial resolution ratio between MS and PAN
N	Number of MS bands

image quality evaluator model, was proposed in [118], and a perceptual quality index for pansharpening was described in [119].

A new FR quality assessment protocol, proposed in [120], is based on the polynomial fitting of multiscale RR quality indexes. A similar approach, recast as a multiscale sequential Bayesian problem, was presented in [121], showing robustness and consistency with qualitative analysis at FR. Finally, an efficient approach based on the combination of both the RR assessment by exploiting Wald's protocol and indexes without reference was proposed in [122].

CONTRIBUTION

Following the successful attempt developed by some of the authors about five years ago in [14], with this article, we move in the same direction, performing a critical new comparison. Similar works can be found in the related literature [6], [12], [14], [17]. In particular, [17] presents a review of the CS, MRA, and VO methods, introducing the idea of meta-analysis for performance assessment. This work does not include a critical comparison, and the review of ML approaches is not provided. Instead, [12] introduces a bird's-eye view of many important contributions specifically dedicated to the topics of pansharpening and resolution enhancement, point cloud data fusion, HS and lidar data fusion, multitemporal data fusion, and big data and social media. This article offers quite broad review of multisource and multitemporal data fusion that is far from the goal of our article. In [6], a review and a critical comparison are performed but for the problem of HS pansharpening, including methods in the CS, MRA, and VO classes. Finally, this article differs from [14], showing a wider comparison involving many recent advances in MS pansharpening. In particular, new developments concerning CS and MRA are included together with methods belonging to the so-called third generation of pansharpening, i.e., VO and ML. Indeed, for the first time, all of these heterogeneous methods are compared in a common framework, fairly assessing both their quantitative and qualitative performance

and the computational burden. The performance assessment is done exploiting both the RR and the FR protocols. An updated version of the QNR index [115], the so-called HQNR [117], is used in this article to assess the performance at FR. HQNR differs from QNR in that it uses a different spectral distortion index, borrowed from Khan's protocol [116]. Finally, to ease the benchmarking and support reproducible research, the implementations of the methods and the framework devoted to the performance assessment have been collected in a MATLAB toolbox that is made available to the community.

NOTATION AND ORGANIZATION

Table 1 reports a list of the main symbols together with a brief description. The other symbols are defined within the article as needed. The notation used in the remainder of the article is detailed in the following. Vectors are indicated in boldface lowercase (e.g., \mathbf{x}) with the i th element indicated as x_i . 2D and 3D arrays are expressed in boldface uppercase (e.g., \mathbf{X}). An MS image $\mathbf{X} = \{\mathbf{X}_k\}_{k=1,\dots,N}$ is a 3D array composed of N bands indexed by the subscript $k = 1, \dots, N$; accordingly, \mathbf{X}_k indicates the k th band of \mathbf{X} . The PAN image is a 2D matrix and will be indicated as \mathbf{P} .

REVIEW OF PANSHARPENING METHODS

This section presents methods belonging to the proposed benchmark. The four classes of pansharpening approaches (CS, MRA, VO, and ML) are briefly described together with several instances belonging to each class.

COMPONENT SUBSTITUTION

The methods belonging to the CS class rely upon the projection of the MS image into a new space, where the spatial structure is well separated from the spectral information [112]. Afterwards, the transformed MS image can be sharpened by replacing the component containing the spatial structure with the PAN image. Finally, the sharpening of the MS image is completed by the inverse transformation applied to the data, bringing them back to the original space. The substitution step usually introduces spectral distortion because of (possible) changes in the low spatial frequencies of the MS image. To alleviate this issue, spectral matching procedures are often applied before substituting the spatial structure of the MS image with the PAN image [14]. It is worth noting that CS-based methods get high fidelity in rendering details in fused products [30]. Furthermore, these methods are usually easy to implement and have a low computational burden, which enables them to be used when a huge amount of data must be fused [14].

Under the hypotheses of linear transformation and the substitution of a sole component, the CS fusion process can be strongly simplified without the explicit calculation of forward and backward transformations [123].

This leads to a faster implementation that can be generally described, for each $k = 1, \dots, N$, by the following equation:

$$\widehat{\mathbf{MS}}_k = \overline{\mathbf{MS}}_k + \mathbf{G}_k \cdot (\mathbf{P} - \mathbf{I}_L), \quad (1)$$

in which the subscript k indicates the k th spectral band; $\mathbf{G}_1, \dots, \mathbf{G}_k, \dots, \mathbf{G}_N$ are the injection gain matrices, typically stacked in a multidimensional array, \mathbf{G} ; and the matrix multiplication is meant pointwise. Finally, \mathbf{I}_L is the so-called intensity component, defined as

$$\mathbf{I}_L = \sum_{i=1}^N w_i \overline{\mathbf{MS}}_i, \quad (2)$$

in which the weight vector $\mathbf{w} = [w_1, \dots, w_i, \dots, w_N]$ is the first row of the forward transformation matrix [112], [124].

Figure 1 presents a flowchart describing the general fusion process for CS-based approaches. The figure points out blocks related to the interpolation, the calculation of the intensity component by (2), the spectral matching between \mathbf{P} and \mathbf{I}_L , and details injected in agreement with (1).

BROVEY TRANSFORM WITH HAZE CORRECTION

Setting the injection gains for each $k = 1, \dots, N$ in (1) as

$$\mathbf{G}_k = \frac{\overline{\mathbf{MS}}_k}{\mathbf{I}_L}, \quad (3)$$

where the division is intended pixel-wise, we have

$$\widehat{\mathbf{MS}}_k = \overline{\mathbf{MS}}_k + \frac{\overline{\mathbf{MS}}_k}{\mathbf{I}_L} \cdot (\mathbf{P} - \mathbf{I}_L) = \overline{\mathbf{MS}}_k \cdot \frac{\mathbf{P}}{\mathbf{I}_L}. \quad (4)$$

This pansharpener approach is the widely known Brovey transform (BT) [125]. Thus, BT fits the general model (1), with injection gains as in (3) representing a sharpening method based on a multiplicative injection scheme, as shown in (4). \mathbf{I}_L can be calculated in several ways. An optimized version relies upon the calculation of \mathbf{I}_L by a weighted average of the MS bands, with weights obtained by minimizing the MSE with respect to a low-pass-filtered version of the PAN image [31].

An interpretation of the multiplicative injection model in (4) in terms of the radiative transfer model leads to the consideration of the path radiance of each MS band, which is undesired energy scattered from the atmosphere [126]. The path radiance, which appears as a haze in a true-color display, should be estimated and subtracted from each band before modulation and reinserted later to restore the unbiased sharpened image. Thus, a haze-corrected version of the BT in (4), called *BT-H*, has recently been proposed [59]. Equation (4) is modified as follows:

$$\begin{aligned} \widehat{\mathbf{MS}}_k &= \overline{\mathbf{MS}}_k + \frac{\overline{\mathbf{MS}}_k - H_k}{\mathbf{I}_L - H_p} \cdot (\mathbf{P} - \mathbf{I}_L) \\ &= \frac{\overline{\mathbf{MS}}_k - H_k}{\mathbf{I}_L - H_p} \cdot (\mathbf{P} - H_p) + H_k, \end{aligned} \quad (5)$$

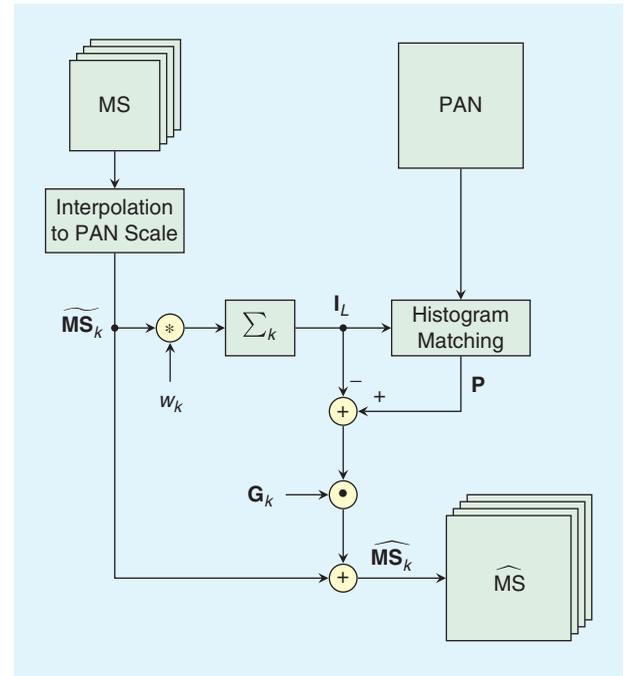


FIGURE 1. A flowchart of pansharpener methods belonging to the CS class.

where H_k denotes the constant haze for the k th MS band and H_p is the haze both of the intensity component and of PAN [59]. For further details about this approach and the estimation of the haze values, interested readers can refer to the original article [59] and the related literature [56], [60].

GRAM-SCHMIDT

The GS orthogonalization procedure has been used for pansharpener [24], taking the name *GS spectral sharpening*. This procedure exploits the intensity component, \mathbf{I}_L , as the first vector of the new orthogonal basis. The orthogonalization processes one MS vector at a time, finding its projection on the (hyper)plane defined by the previously found orthogonal vectors and its orthogonal component such that the sum of the orthogonal and projection components is equal to the zero-mean version of the original vectorized band. Pansharpener is completed by substituting \mathbf{I}_L with the histogram-matched \mathbf{P} before performing the inverse transformation.

The procedure can be recast into the flowchart in Figure 1. Thus, the fusion process is described by (1) with the following injection gains for $k = 1, \dots, N$ [30]:

$$\mathbf{G}_k = g_k \mathbf{1} = \frac{\text{cov}(\overline{\mathbf{MS}}_k, \mathbf{I}_L)}{\text{var}(\mathbf{I}_L)} \mathbf{1}, \quad (6)$$

in which $\mathbf{1}$ is an all-one matrix; $\text{cov}(\mathbf{X}, \mathbf{Y})$ indicates the covariance between two images, \mathbf{X} and \mathbf{Y} ; and $\text{var}(\mathbf{X})$ is the variance of \mathbf{X} .

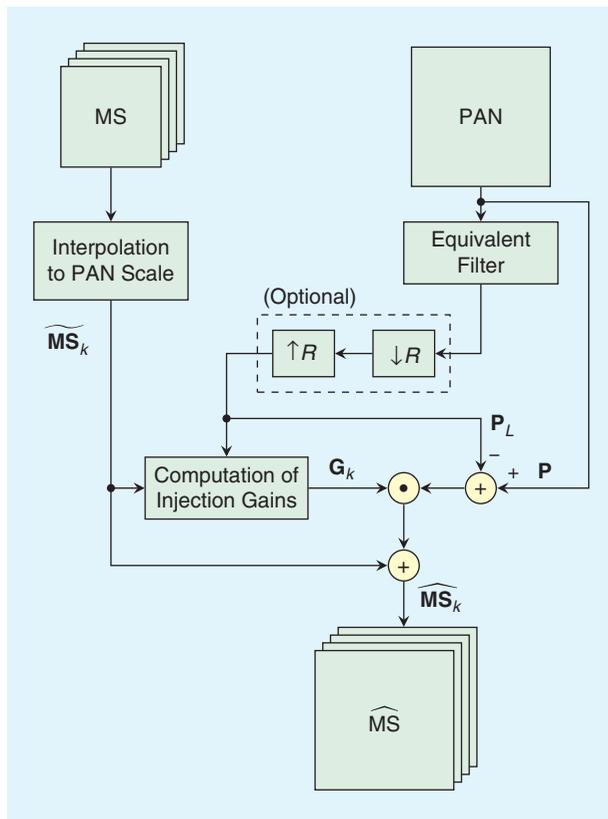


FIGURE 2. A flowchart of pansharpening methods belonging to the MRA class. Some MRA approaches skip the dashed box.

Several versions of GS are achieved by changing the way to generate I_L : by averaging the MS components (i.e., $w_i = 1/N$, for all $i = 1, \dots, N$) in the original GS [24]; by a weighted average of the MS bands minimizing the MSE with respect to a low-pass-filtered version of a PAN image in the GSA [30] and by applying GSA separately to each cluster obtained through k -means clustering applied to the MS image in the context-adaptive GSA (C-GSA) [32].

BAND-DEPENDENT SPATIAL-DETAIL-BASED APPROACHES

The BDSB-based approaches [33] start from an extended version of the generic formulation (1) for all $k = 1, \dots, N$. Thus, we have

$$\widehat{MS}_k = \overline{MS}_k + g_k \left(P - \sum_{i=1}^N w_{k,i} \overline{MS}_i \right). \quad (7)$$

The optimal minimum MSE joint estimation of the weights, $w_{k,i}$, and scalar gains, g_k , in (7) would encompass the use of the unknown target image, \widehat{MS}_k . Thus, the problem is solved at RR by exploiting the original MS image as a reference image [33]. A context-adaptive BDSB (C-BDSB) has also been proposed in [34]. A physically constrained optimization (BDSB-PC), improving the quality of the fusion product, particularly for eight-band MS data sets, has been recently proposed in [52].

PARTIAL REPLACEMENT ADAPTIVE COMPONENT SUBSTITUTION

The concept of partial replacement of the intensity component is introduced in [29]. The PAN image is not directly used for CS. Instead, the algorithm uses a weighted sum of the PAN and the k th MS band to calculate the k th sharpened band in (1). For this reason, this method is referred to as PRACS.

MRA

MRA methods extract the PAN details exploiting the difference between P and its low-pass spatial version, P_L . Namely, the fused image is obtained for each $k = 1, \dots, N$ as follows:

$$\widehat{MS}_k = \overline{MS}_k + G_k \cdot (P - P_L). \quad (8)$$

The different approaches belonging to this class are characterized by the way to calculate P_L and by the employed injection gains $\{G_k\}_{k=1, \dots, N}$. In a very general setting, P_L is achieved through an iterative decomposition scheme (called MRA) that aims at constructing a sequence of 2D signals with progressively reduced information through the repeated application of some analysis operators.

The general fusion scheme for MRA approaches is displayed in Figure 2. We can note blocks devoted to the interpolation of the MS image to reach the PAN scale; the calculation of the low-pass version P_L of the PAN image depending on the resolution ratio, R ; the computation of the band dependent injection gains, $\{G_k\}_{k=1, \dots, N}$; and the fusion according to (8). It is worth noting that there is a difference if P_L is decimated-interpolated (dashed box) or not. The first solution is usually much more appreciated due to the possibility of compensating for the aliasing of the MS image through the fusion process [127]. Finally, MRA-based fused products reduce the spectral distortion but often compromise the spatial rendering. They are also robust to temporal misalignments, which is a valuable feature when multiplatform fusion is considered [128].

ADDITIVE WAVELET LUMINANCE PROPORTIONAL

The undecimated “à trous” method [44] has recently emerged as a very effective one for extracting PAN details [129]. Indeed, even if nonorthogonality could compromise the spectral quality of fused products [129], its beneficial characteristics, such as the shift-invariance property [130] and the capability of being easily matched to the sensor MTF [47], produce accurate pansharpened images. The AWLP, originally proposed by [61], exploits the “à trous” wavelet for details of extraction from the PAN image using the fusion rule in (8) and injection coefficients estimated as

$$G_k = \frac{\overline{MS}_k}{\frac{1}{N} \sum_{i=1}^N \overline{MS}_i}. \quad (9)$$

To get high performance, a preliminary histogram-matching procedure is also required [31].

GENERALIZED LAPLACIAN PYRAMID

P_L at the original MS resolution can be performed with multiple fractional steps. This method is commonly referred to as *pyramidal decomposition* and dates back to the seminal work of Burt and Adelson [36], which utilized Gaussian low-pass filters to carry out the analysis steps. The corresponding differential representation, achieved by calculating the differences between the Gaussian pyramid levels, is named the Laplacian pyramid and later was proved to be very valuable for pansharpening purposes [43]. In [48], the authors showed that a single Gaussian low-pass filter with a cut frequency equal to $1/R$ (where R is the resolution ratio between PAN and MS) and decimating by R is enough to get high performance. Gaussian filters can be tuned to closely match the MS sensor's MTF. This allows extracting from the PAN image those details that are not seen by the MS sensor [47]. Since the Gaussian mask is defined by a single parameter (i.e., its standard deviation), its frequency response is fully specified by fixing that parameter. To this aim, the value of the amplitude response at the Nyquist frequency is used, since it is commonly provided by the manufacturer.

In the literature, many instances of generalized Laplacian pyramid (GLP) approaches, relying upon filters that exploit the MS sensor's MTF, have been proposed by changing the way to estimate the injection coefficients. High performance is usually expected by these methods, motivating their inclusion into the proposed benchmark. Their brief description is provided in the following:

- The classical MTF-GLP is based on MTF Gaussian filters for detail extraction and an additive injection model, i.e., $G_k = \mathbf{1}$ for each $k = 1, \dots, N$. To get high performance, a histogram matching procedure between P_L and each MS band is required, as proposed in [31].
- MTF-GLP-HPM, where HPM stands for high-pass modulation, adopts the injection model exploited in the BT (3) by substituting P_L with I_L , i.e.,

$$G_k = \frac{\overline{MS}_k}{P_L}, \quad (10)$$

where the division is intended as pixel-wise. Even in this case, a preliminary histogram matching is required [31]. Improved versions are 1) MTF-GLP-HPM-R [131], adopting a spectral matching procedure based on the multivariate linear regression between each MS band and a low-pass version of the PAN image; and 2) MTF-GLP-HPM-H [59], a haze-corrected version like the one in BT-H (see the "Brovey Transform With Haze Correction" section), whose fusion rule is modified as follows:

$$\widehat{MS}_k = \overline{MS}_k + \frac{\overline{MS}_k - H_k}{P_L - H_p} \cdot (P - P_L). \quad (11)$$

- MTF-GLP-CBD (i.e., MTF-GLP with a context-based

decision) adopts a projective, or regression-based, injection model [47], [132] in conjunction with the MTF-GLP framework. The injection coefficients are estimated through multivariate linear regression on \overline{MS}_k and P_L ,

$$G_k = g_k \mathbf{1} = \frac{\text{cov}(\overline{MS}_k, P_L)}{\text{var}(P_L)} \mathbf{1}, \quad (12)$$

in a way similar to that for the GS-based approaches (see the "Gram-Schmidt" section) but considering P_L instead of I_L . A local version based on the use of a k -means clustering to segment the MS image in spectrally coherent parts is also proposed in the literature and here named context-adaptive MTF-GLP-CBD (C-MTF-GLP-CBD) [32].

- MTF-GLP-FS presents a new fusion rule at full scale (FS) considering a projective injection scheme that removes the hypothesis of invariance among scales [133]. Its asymptotic solution leads to

$$G_k = g_k \mathbf{1} = \frac{\text{cov}(\overline{MS}_k, P)}{\text{cov}(P_L, P)} \mathbf{1}. \quad (13)$$

MORPHOLOGICAL FILTERS

A nonlinear decomposition scheme has recently been investigated in the MS pansharpening literature [55]. The nonlinear MRA scheme was implemented with a morphological pyramid based on morphological half gradients. The approach can be recast into the general MRA fusion scheme (8) using an HPM injection model [see (10)]. Due to the use of MF for the detail extraction phase, this method, proposed in [55] and introduced in our benchmark, will be called *MF*.

VARIATIONAL OPTIMIZATION-BASED TECHNIQUES

In this section, we present some examples of techniques that belong to the VO category, as they rely on an acquisition or representation model leading to a variational optimization problem.

FILTER ESTIMATION

The article [50] presents a variational approach for estimating a spatial degradation filter that links the full- and reduced-scale spatial resolution acquisitions. This spatial filter is often associated with the MTF of the MS sensor, whose value at the Nyquist frequency is often available. However, if this information is not available or not precise, the filter \mathbf{h} should be estimated. This can be done as in [50] by relying on the model where a low-resolution PAN can be approximated to a linear combination of the MS bands. The corresponding variational problem can be cast as

$$\hat{\mathbf{h}} = \underset{\mathbf{h}, \mathbf{w}}{\text{argmin}} \{ \|\mathbf{P}_C \mathbf{h} - \mathbf{Y} \mathbf{w}^T\|^2 + \lambda \mathcal{R}(\mathbf{h}) \}, \quad (14)$$

where \mathbf{Y} is the MS image interpolated to the PAN scale, i.e., with r pixels, represented in matrix format ($\mathbf{Y} \in \mathbb{R}^{r \times N}$); $\mathbf{P}_C \mathbf{h}$ represents the linear convolution in matrix form between the

PAN image \mathbf{p} (lexicographically ordered) and the blur matrix \mathbf{h} ; \mathbf{w} is the row vector of the weights to project MS into the PAN domain; and \cdot^T indicates the transpose operator. The degradation filter h is represented by block matrix \mathbf{h} , and $\mathcal{R}(\mathbf{h})$, defined in [50], is a regularization term on \mathbf{h} , weighted by the coefficient λ . Once the degradation filter has been estimated, it can be used in conventional MRA approaches.

SPARSE REPRESENTATION OF INJECTED DETAILS

The pansharpening technique presented in [88] is an example of methods based on sparse representation. In [88], the authors propose to generate the spatial details, which will be injected by a conventional MRA scheme (8), using a dictionary of patches. Specifically, the dictionary \mathbf{D}^h at FS is composed of patches representing high-spatial-resolution details. As is common in dictionary-based pansharpening approaches, the coefficients $\boldsymbol{\alpha}$ of the linear combination are estimated by solving a variational problem (i.e., a sparse regression here) at the native spatial resolution of the MS image. Working at RR allows considering a reference image in the coefficient estimation. Thus, scale invariance is a strong hypothesis for such methods. The variational problem to solve is in the form

$$\hat{\boldsymbol{\alpha}} = \operatorname{argmin} \|\boldsymbol{\alpha}\|_0 \quad \text{such that} \quad \mathbf{y} = \mathbf{D}^h \boldsymbol{\alpha}, \quad (15)$$

with \mathbf{y} being a patch and \mathbf{D}^h a dictionary of details at a reduced scale. The sparsity penalty induced by the ℓ_0 pseudonorm on the coefficients enforces a parsimonious representation of the details from a few patches. The estimated coefficients will be used for the representation of the FS details (i.e., $\mathbf{y}^h = \mathbf{D}^h \boldsymbol{\alpha}$). This paradigm effectively exploits the self-similarity of images and their parsimonious representations, which are particularly valid for an image of spatial details.

TOTAL VARIATION PANSHARPENING

The cost function for the TV pansharpening method in [67] is given by the following TV-regularized least squares problem:

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|^2 + \lambda \mathbf{TV}(\mathbf{x}), \quad (16)$$

where $\mathbf{y} = [\mathbf{y}_{\text{MS}}^T, \mathbf{y}_{\text{PAN}}^T]^T$, \mathbf{y}_{MS} , and \mathbf{y}_{PAN} are the MS in matrix format and the PAN in vector, $\mathbf{M} = [\mathbf{M}_1^T, \mathbf{M}_2^T]^T$, \mathbf{M}_1 is a decimation matrix, \mathbf{M}_2 reflects that the PAN image is assumed to be a linear combination of the MS bands, λ is a weight, and $\mathbf{TV}(\cdot)$ is an isotropic TV regularizer. The pansharpened image \mathbf{x} is found by minimizing the convex cost function (16). However, the minimum does not have a closed form solution. Fortunately, it can efficiently be found by using the majorization–minimization algorithm detailed in [67].

PRINCIPAL COMPONENT ANALYSIS/WAVELET MODEL-BASED FUSION

The PCA/wavelet model-based fusion (PWMBF) method [79] assumes the following model:

$$\mathbf{Y} = \mathbf{W}\mathbf{Z} + \mathbf{N}, \quad (17)$$

where \mathbf{Y} contains the vectorized observed MS bands in its columns, \mathbf{W} is a degradation (blurring and decimation) operator, \mathbf{Z} is the vectorized pansharpened MS bands in its columns, and \mathbf{N} is noise. It is assumed that \mathbf{Z} and the PAN image are jointly Gaussian. The main assumption behind this method is that \mathbf{Y} and \mathbf{Z} share PC so that the model can be rewritten as

$$\mathbf{B} = \mathbf{W}\mathbf{G} + \mathbf{N}, \quad (18)$$

where \mathbf{B} contains the PC scores corresponding to \mathbf{Y} and \mathbf{G} is an $r \times q$ matrix corresponding to \mathbf{Z} , with r being the number of pixels of the PAN image and q the number of PCs used in the model. The pansharpening is based on solving the maximum *a posteriori* problem

$$\hat{\mathbf{g}} = \operatorname{argmax} p(\mathbf{g}|\mathbf{b}, \mathbf{y}_{\text{PAN}}), \quad (19)$$

where \mathbf{g} and \mathbf{b} represent the wavelet-transformed and wavelet-vectorized matrices \mathbf{G} and \mathbf{B} . By making certain assumptions, such as the spatial independence of wavelet coefficients, an efficient algorithm can be derived. This algorithm is detailed in [79].

REDUCED-RANK-BASED PANSHARPENING

The main idea behind reduced-rank pansharpening is to exploit the spectral correlations between the MS and PAN images. This is done by defining a matrix, \mathbf{X} , that has the PAN image in its first column and the pansharpened MS images in column 2 to $N + 1$. We then assume that it has a low-rank representation,

$$\mathbf{X} = [\mathbf{x}_i]_{i=1}^{N+1} = \mathbf{G}\mathbf{F}^T, \quad (20)$$

where $\mathbf{G} = [\mathbf{g}_j]_{j=1}^l$, $\mathbf{F} = [\mathbf{f}_i^T]_{i=1}^l$ is an orthonormal matrix and $l < N + 1$ is the low rank. The individual bands can be retrieved from this low-rank representation as $\mathbf{x}_i = \mathbf{G}\mathbf{f}_i$. The pansharpening problem is then to minimize the following cost function:

$$J(\mathbf{F}, \mathbf{G}) = \sum_{i=1}^{N+1} \frac{1}{2} \|\mathbf{y}_i - \mathbf{M}_i \mathbf{B}_i \mathbf{G} \mathbf{f}_i\|^2 + \sum_{j=1}^l \lambda_j \phi_w(\mathbf{g}_j), \quad (21)$$

where \mathbf{B}_i and \mathbf{M}_i are the blurring and decimation matrix corresponding to band i , respectively; $\lambda_j > 0$ are the weights; and $\phi_w(\mathbf{g}_j)$ is a weighted roughness penalty further described in [134]. The cost function is minimized by using a cyclic descent algorithm that alternates between minimizing with respect to \mathbf{F} and \mathbf{G} . The optimization problem with respect to \mathbf{G} is a convex problem and is solved efficiently by using conjugate gradient. The orthonormality-constrained optimization problem with respect to \mathbf{F} is efficiently solved by using manifold optimization [135], [136]. Further details are given in [72], [134].

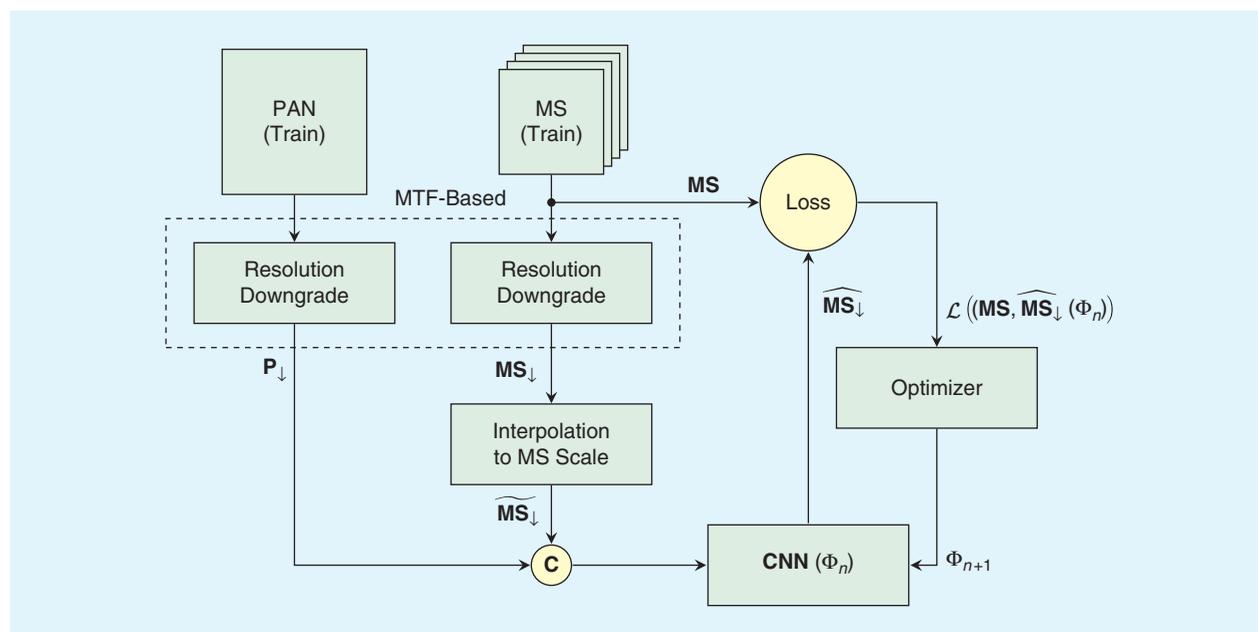


FIGURE 3. The training scheme for PNN. The dashed box encloses the resolution downgrading process (similar to Wald’s protocol). At each iteration, a different MS-PAN pair (or minibatch of pairs) is used to update the CNN parameters. **C**: concatenation.

MACHINE LEARNING

Nowadays, among ML approaches, CNN-based methods are the most promising ones. Therefore, we focus here on a family of solutions rooted in the pioneering work of Masi et al. [91], among the first to use CNNs for pansharpening. In particular, we present four models, two of which, hereinafter referred to as *PNN* and *PNN-IDX* (where *IDX* stands for the use of some indexes as auxiliary inputs), were introduced in [91], and the other two, advanced versions of *PNN*, here named *A-PNN* and *A-PNN-FT* (where *FT* stands for fine-tuning), were proposed more recently in [95].

The development of a DL method for pansharpening involves at least the following steps:

- 1) training data set setup
- 2) network architecture design
- 3) definition of a proper loss function to guide the training
- 4) network training using a proper optimization method
- 5) network validation on a dedicated data set.

As a peculiar trait, all CNN-based methods for pansharpening share a common key problem, which is related to the collection of data for training (step 1), i.e., a sufficiently “rich” set of input-output examples to learn the unknown network parameters (say, Φ). Ideally, pansharpened images are, in fact, unavailable and therefore must be synthesized using an ad hoc generation procedure. Figure 3 depicts the conceptual scheme of the training process proposed in [91]. The underlying idea is a resolution shift, such that the MS component plays the role of reference output rather than input. This is achieved by properly downgrading the resolution (dashed section) of both MS and PAN components of any training image with a resolution ratio of R . Then, an interpolation of the RR MS (MS_{\downarrow}) back to its original size provides a properly

sized component ($\widehat{MS}_{\downarrow}$) to be concatenated with the downgraded PAN image (P_{\downarrow}) prior to feeding the CNN. In practice, this process is repeated on one or more images reserved for training and combined with a cropping process. By doing so, we come to have thousands of input-output pairs in an RR domain. This allows organizing the training data in minibatches for efficient iterative optimization. In particular, at iteration n the pansharpening outcome $\widehat{MS}_{\downarrow}$ is compared to the corresponding reference (original MS) using a suitably defined loss function, $\mathcal{L}(\cdot, \cdot)$. Taking the gradient of the loss, eventually the optimizer computes the updated CNN parameter set Φ_{n+1} . All models presented here make use of a stochastic gradient descent algorithm with momentum. The resolution downgrading process, carried out according to the sensor MTF, soon became a standard option for CNN-based pansharpening methods after its introduction in [91], to which interested readers are referred for further details. Once the training process is completed, the network is ready to be used on test images, as summarized in Figure 4 for the two variants proposed in [91], *PNN* and *PNN-IDX* (including the dashed section and by properly training the CNN, also considering these new inputs).

Network architecture design (step 2) is probably the item that most characterizes the different proposals. In general terms, it amounts to a definition of a directed acyclic graph (DAG), which describes the input-output information flow, associating a specific task (such as convolution, point-wise nonlinearities [e.g., rectified linear unit (ReLU)], batch normalization, concatenation, sum, and pooling) to each DAG vertex. Specific subgraph structures obtained by combining these elementary operations are also often employed (e.g., residual, dense, or inception modules). The *PNN* model

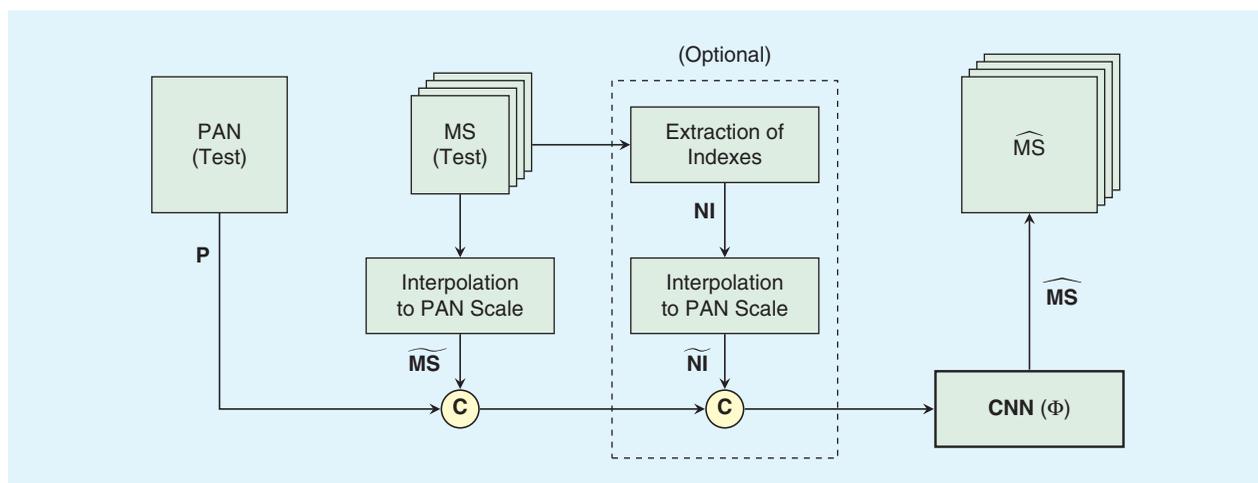


FIGURE 4. The top-level flowchart of the PNN and PNN-IDX (dashed box included) algorithms. NI is a stack of normalized indexes extracted from MS.

TABLE 2. THE CNN MODEL HYPERPARAMETERS.

CONVOLUTIONAL LAYER	SENSOR	MODEL	SPATIAL SUPPORT	INPUT BANDS	OUTPUT BANDS
1	IK	PNN, A-PNN/-FT	5 × 5	5	48
		PNN-IDX	5 × 5	7	48
	GE-1	PNN, A-PNN/-FT	9 × 9	5	48
		PNN-IDX	9 × 9	7	48
WV-3	PNN, A-PNN/-FT	9 × 9	9	48	
	PNN-IDX	9 × 9	13	48	
2	all	all	5 × 5	48	32
3	IK, GE-1	all	5 × 5	32	4
	WV-3	all	5 × 5	32	8

[91] is a relatively simple one: a serial net composed of three convolutional units interleaved by ReLU activations. The main architectural hyperparameters of the PNN are gathered in Table 2.

The PNN-IDX variant of PNN involves the use of auxiliary inputs derived from the MS bands. These are well-established radiometric indexes such as the normalized difference vegetation/water indexes, considered for all data sets of the present work, or the normalized difference soil index and nonhomogeneous feature difference, which are used only for WV-3 images. Therefore, the main difference between PNN and PNN-IDX concerns the CNN input formation process, highlighted in Figure 4 with a dashed box, which implies a larger number of input bands, as results from the model hyperparameters reported in Table 2. The use of these additional input bands was motivated by the experimental observation that many feature maps extracted by the first convolutional layer of a PNN have shown a strong correlation with these radiometric indexes. The intuition was then confirmed experimentally with a superior performance of PNN-IDX over PNN.

From the architectural perspective, one of the most important innovations registered in recent years in the design of CNNs was the introduction of residual learning schemes [107], whose primary purpose was speeding up training. Residual learning was soon demonstrated to be a natural choice for resolution enhancement [137]. In fact, the desired superresolved image can be viewed as composed of its low- and high-frequency components, the former being essentially the input low-resolution image and the latter being the missing (or residual) part to be actually restored. Residual schemes naturally address superresolution or pansharpening problems in light of this partition, avoiding the unnecessary reconstruction of the whole desired output and reducing the risk of altering the low-frequency content of the image. As a matter of fact, most recent DL pansharpening models, including the advanced versions of PNN (A-PNN/-FT) [95], embed residual modules [92], [94], [95], [97], [99], [100], although they can be very different in terms of complexity. In addition to this, in [95], a different loss function, L_1 -norm, was also used, proving to be more effective than the L_2 -norm used in [91]. Finally, a target-adaptive FT phase has also been introduced. Here, we will consider the two variants with or without FT, A-PNN-FT and A-PNN, respectively. The full version is summarized in Figure 5. Solid-line connections refer to the FT phase. The same scheme, but with input images fetched from the training data set, is used for the network pretraining, which is supposed to have been run prior to entering the FT. The network parameters of A-PNN will, therefore, represent the starting point for A-PNN-FT when the FT starts. The residual learning configuration is easily recognizable by comparing the training flowchart in Figure 5 (solid lines) with that of a PNN depicted in Figure 3. In fact, the only difference is the presence of a skip connection in the former that brings the upscaled MS component directly to the CNN output to be added to the network outcome (detail component). The CNN architecture actually remains the same (see hyperparameters of A-PNN/-FT versus PNN in Table 2), although

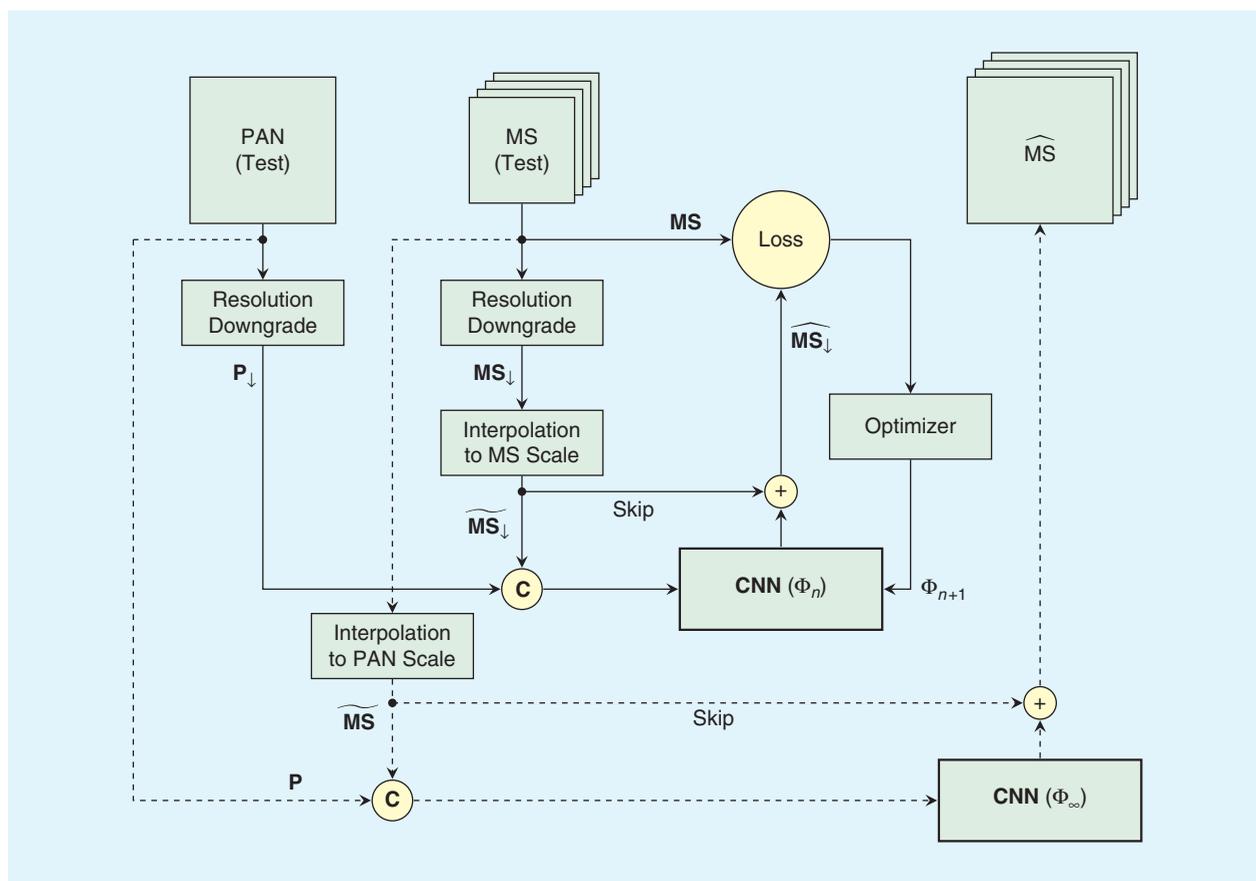


FIGURE 5. The A-PNN-FT top-level workflow. Solid connections correspond to the iterative FT process. Upon convergence, the trained parameters Φ_{∞} are used at the inference time (dashed connections) to provide the FR pansharpener.

the trained parameters will eventually be different. Once the FT is concluded (see details in [95]), the network parameters are frozen (say, Φ_{∞}) and used to infer the pansharpener of the FR target image (dashed lines) skipping the resolution downgrade.

QUALITY ASSESSMENT OF FUSION PRODUCTS

Quality assessment of pansharpener methods and data products has been extensively investigated in the past two decades. After a first generation of empirical methods matching subjective evaluation, an objective definition of quality and a related measurement protocol were introduced [110].

According to this protocol, fusion products should satisfy two main properties: consistency and synthesis. Consistency indicates that the pansharpener image, degraded at the original MS resolution, should be as similar as possible (spectrally) to the original MS image. Synthesis requires that the pansharpener image should be similar to the one acquired by a greater resolution MS sensor. Verifying the synthesis property is possible by performing fusion with both MS and PAN data sets degraded at spatial resolutions lower than those of the original ones.

While synthesis is extensively adopted by following an RR approach, checking consistency at FR is more critical,

since the measured spectral and spatial qualities may follow opposite trends, with the paradox that the least spectrally distorted fused image is that obtained when no spatial enhancement is introduced.

REDUCED RESOLUTION ASSESSMENT

RR assessment measures the similarity of the fused product to an ideal reference, i.e., the original MS image. That is possible by degrading the resolutions of both the original MS and the original PAN and by performing fusion from those degraded data. Clearly, the choice of the filter is crucial in this validation protocol. In general, the filter is defined to ensure the consistency property of the pansharpener process. Thus, it is straightforward that the resolution reduction of the MS image has to be done by exploiting spatial filters matching the MS sensor's MTFs [47]. In addition, the filter used to degrade the PAN image must be designed to preserve the details that would have been seen if the image were acquired at the RR. Accordingly, a common choice is the use of an ideal filter [47].

The more similar the obtained pansharpener image is to the original MS image, the higher the measured quality. Such a similarity degree can be easily computed through score indexes that compare two multiband images.

SPECTRAL ANGLE MAPPER

The spectral angle mapper (SAM) was originally introduced for discrimination of materials starting from their reflectance spectra [138]. Given two spectral vectors, \mathbf{v} and $\hat{\mathbf{v}}$, both having N components, in which $\mathbf{v} = [v_1, v_2, \dots, v_N]$ is the reference spectral pixel vector and $\hat{\mathbf{v}} = [\hat{v}_1, \hat{v}_2, \dots, \hat{v}_N]$ is the test spectral pixel vector, SAM denotes the absolute value of the spectral angle between the two vectors as

$$\text{SAM}(\mathbf{v}, \hat{\mathbf{v}}) = \arccos\left(\frac{\langle \mathbf{v}, \hat{\mathbf{v}} \rangle}{\|\mathbf{v}\|_2 \cdot \|\hat{\mathbf{v}}\|_2}\right). \quad (22)$$

SAM is usually expressed in degrees and is equal to zero if and only if the test vector is spectrally identical to the reference vector; i.e., the two vectors are parallel and may differ only by their moduli. A global spectral dissimilarity, or distortion, index is obtained by averaging (22) over the whole scene.

ERREUR RELATIVE GLOBALE ADIMENSIONNELLE DE SYNTHÈSE

The ERGAS index, whose French acronym stands for *relative dimensionless global error in synthesis* [139], is a normalized dissimilarity index that offers a global indication of the distortion toward the reference of a test multiband image:

$$\text{ERGAS} = 100 \frac{d_h}{d_l} \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{\text{RMSE}(n)}{\mu(n)} \right)^2}, \quad (23)$$

where d_h/d_l is the ratio between the pixel sizes of PAN and MS, typically 1/4 for many sensors used for pansharpening; $\mu(l)$ is the mean (average) of the n th band of the reference; and N is the number of bands. Low values of ERGAS indicate high similarity between fused and reference MS data.

$Q2^n$

$Q2^n$ is the multiband extension of the universal image quality index (UIQI) [140] and was introduced for quality assessment of pansharpened MS images, first for four bands [141] and later extended to 2^n bands [142]. Each pixel of an image with N spectral bands is accommodated into a hypercomplex (HC) number with one real part and $N-1$ imaginary parts.

Let $\mathbf{z} = \mathbf{z}(m, n)$ and $\hat{\mathbf{z}} = \hat{\mathbf{z}}(m, n)$ denote the HC representation of the reference and test spectral vectors at pixel (m, n) . Analogously to UIQI, namely, $Q2^0 = Q$, $Q2^n$ may be written as the product of three terms:

$$Q2^n = \frac{|\sigma_{z\hat{z}}|}{\sigma_z \sigma_{\hat{z}}} \cdot \frac{2\sigma_z \sigma_{\hat{z}}}{\sigma_z^2 + \sigma_{\hat{z}}^2} \cdot \frac{2|\bar{z}| |\bar{\hat{z}}|}{|\bar{z}|^2 + |\bar{\hat{z}}|^2}, \quad (24)$$

the first of which is the modulus of the HC correlation coefficient between z and \hat{z} . The second and the third terms measure contrast changes and mean bias, respectively, on all bands simultaneously. Statistics are calculated on $N \times N$ blocks, typically, 32×32 , and $Q2^n$ is averaged over the blocks of the whole image to yield the global score index. $Q2^n$ takes values in $[0, 1]$ and is equal to one if and only if $z = \hat{z}$ for all pixels.

FULL-RESOLUTION ASSESSMENT

FR assessment infers the quality of the pansharpened image at the scale of the PAN image without resorting to a single reference image, which is not available [117], [120]. Consequently, the problem of assessing the quality of pansharpened products at FR is intrinsically ill posed. To solve the problem, new distortion measurements have been introduced [115], [116], such that they do not depend on the unavailable true HRMS data but would measure zero distortions if such data were available. Two score indexes for FR assessment are reported in the following.

QUALITY WITH NO REFERENCE

The QNR protocol takes separate measurements, each normalized in $[0, 1]$ of spectral and spatial consistencies [115]. The motivation underlying this protocol is that the mutual similarities between any couples of MS bands and each MS band and PAN should be unchanged on average, varying with the resolution, that is, before and after fusion.

The spectral distortion, D_λ , is calculated between the LRMS images and the fused MS images [115]. Two sets of interband UIQI values are calculated separately at low and high resolutions. The differences of corresponding UIQI values at the two scales yield the spectral distortion introduced by the pansharpening process.

The spatial distortion, D_s , combines the UIQI values computed between each MS band and the PAN image degraded to the resolution of MS and again between fused MS and FR PAN. The absolute difference, averaged over all bands, between the corresponding UIQI values yields the spatial distortion D_s [115]. The spatial alignment between the original MS data and the PAN image is of crucial importance to have an accurate evaluation of this distortion index. The latter issue is considered to be solved by the upsampling of the MS image to the PAN scale. Thus, the implementation of this index can be performed considering the problem at the PAN scale, i.e., exploiting the upsampled MS image instead of the original LRMS image.

The two distortions are then combined to provide a unique quality index, referred to as $\text{QNR} \in [0, 1]$, with 1 being the best attainable value:

$$\text{QNR} = (1 - D_\lambda)^\alpha (1 - D_s)^\beta, \quad (25)$$

where usually $\alpha = \beta = 1$.

HYBRID QUALITY WITH NO REFERENCE

The HQNR [117] borrows the spatial distortion index D_s from QNR and the spectral distortion index from Khan's protocol [116], i.e., $D_\lambda^{(K)} = 1 - Q2^n(\widehat{\text{MS}}_i, \overline{\text{MS}})$, where $\widehat{\text{MS}}_i$ is the MTF-filtered pansharpened MS image considering a resolution ratio equal to R and $\overline{\text{MS}}$ is the original MS image interpolated to the PAN scale, R times lower than the MS scale. Again, this implementation is performed by exploiting the upsampled MS image, i.e., working at the PAN scale.

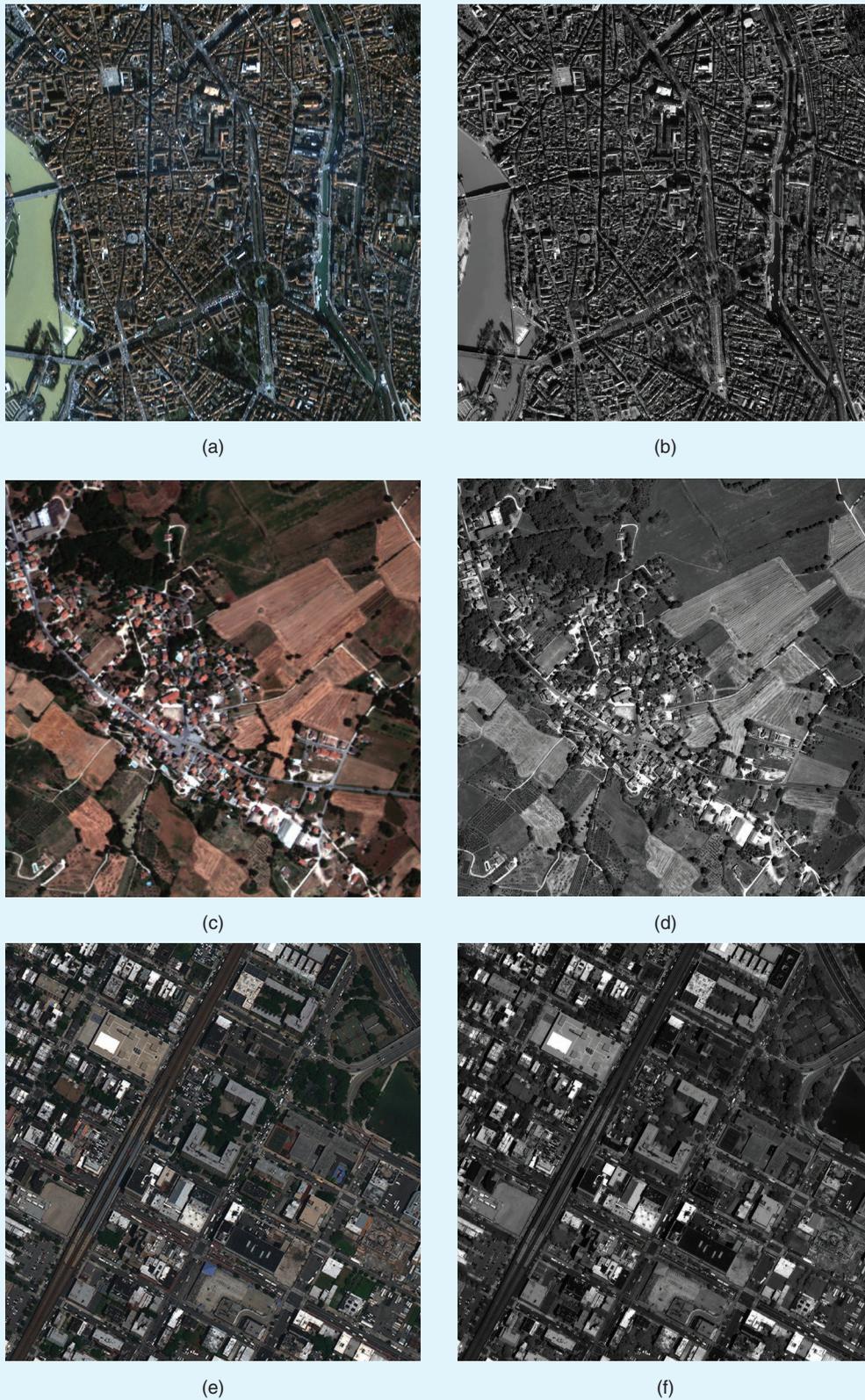


FIGURE 6. The data sets. (a) and (b) Toulouse data set, (c) and (d) Collazzone data set, and (e) and (f) NY data set. The first column contains the MS images upsampled to the PAN scale, and the second column is about the PAN images.

It should be noted that $D_{\lambda}^{(k)}$ strictly follows the consistency property and avoids possible drawbacks of D_{λ} , which computes similarity indexes at different resolutions.

The two distortions are finally combined to yield the unique quality index

$$\text{HQNR} = (1 - D_{\lambda}^{(k)})^{\alpha} (1 - D_s)^{\beta}, \quad (26)$$

where usually $\alpha = \beta = 1$.

EXPERIMENTAL RESULTS

This section is devoted to the presentation of the experimental results.

DATA SETS

Three different scenarios acquired by three different sensors are considered in this article. For all of the cases, both the assessment at RR and at FR, following the indications drawn in the "Quality Assessment of Fusion Products" section, are provided, yielding six different test cases. The characteristics of the employed data sets are detailed in the following.

TOULOUSE DATA SET

This data set, exhibited in Figure 6(a) and (b), represents an urban area of the city of Toulouse (France). It is acquired by the *IK* sensor, which works in the visible and near-infrared spectrum range. The MS sensor is characterized by four bands (blue, green, red, and near infrared), and also a PAN channel is available. The resolution cell is 4 m × 4 m for the MS bands and 1 m × 1 m for the PAN channel. The resolution ratio R is, therefore, equal to four. The radiometric resolution is 11 bits. The size of an MS spectral band is 512 × 512 pixels. Thus, the PAN image is 2,048 × 2,048 pixels. For the test case at RR, both MS and PAN images are properly degraded, in agreement with Wald's protocol. The ground-truth (GT) image is represented by the original MS image, and the low-pass version of the MS image has a size equal to 128 × 128 pixels.

COLLAZZONE DATA SET

This data set depicted in Figure 6(c) and (d) refers to a rural area of Collazzone, a small town in Central Italy. It is acquired by the *GE-1* sensor, which works in the visible and near-infrared spectrum range. The MS sensor is characterized by four bands (blue, green, red, and near infrared), and also a PAN channel is available. The spatial sampling interval (SSI) is 2 m for MS and 0.5 m for PAN, respectively. Thus, R is equal to four. The radiometric resolution is 11 bits. The size of an MS spectral band is 512 × 512 pixels. Therefore, the PAN image is 2,048 × 2,048 pixels. For the test case at RR, both MS and PAN images are properly degraded, in agreement with Wald's protocol as for the Toulouse data set.

NEW YORK DATA SET

The New York (NY) data set, provided in Figure 6(e) and (f), represents an urban area of the city of New York with a more relevant presence of high buildings with respect to European

urban scenarios, such as the one captured in the Toulouse data set. These data are acquired by the (*WV-3*) sensor, which works in the visible and near-infrared spectrum range. The MS sensor is characterized by eight spectral bands (coastal, blue, green, yellow, red, red edge, NIR1, and NIR2), and also a PAN channel is available. The SSI is 1.2 m for MS and 0.3 m for PAN, respectively. Therefore, R is equal to four. The radiometric resolution is 11 bits. The size of an MS spectral band is 512 × 512 pixels. Thus, the PAN image is 2,048 × 2,048 pixels. For the test case at RR, both MS and PAN images are properly degraded, in agreement with Wald's protocol, as for the Toulouse data set.

BENCHMARK

Several state-of-the-art algorithms are employed for comparison purposes:

- ▶ EXP: MS image interpolation using a polynomial kernel with 23 coefficients [43]
- ▶ CS methods
 - BT-H [59]
 - BDSO [33]
 - C-BDSO [34]
 - BDSO-PC [35]
 - GS [24]
 - GSA [30]
 - C-GSA [32]
 - PRACS [29]
- ▶ MRA methods
 - AWLP [61] with revised statistical matching between PAN and MS bands [31]
 - MTF-GLP: GLP [43] with MTF-matched filter [47], unitary injection model, and revised statistical matching between PAN and MS bands [31]
 - MTF-GLP-FS: GLP [43] with MTF-matched filter [47] with a new FS regression-based injection model [133]
 - MTF-GLP-HPM: GLP with MTF-matched filter [47] with HPM injection model [143] and revised statistical matching between PAN and MS bands [31]
 - MTF-GLP-HPM-H: GLP with MTF-matched filter [47] with HPM injection model [143] and haze correction [59]
 - MTF-GLP-HPM-R: GLP with MTF-matched filter [47] and HPM injection model [143] with a new preliminary regression-based spectral matching phase [131]
 - MTF-GLP-CBD: GLP [43] with MTF-matched filter [47] and regression-based injection model [132]
 - C-MTF-GLP-CBD: context-based GLP [43] with MTF-matched filter [47] and regression-based injection model [132] with local parameter estimation exploiting clustering [32]
 - MF: nonlinear decomposition scheme using MFs based on half gradient [55]
- ▶ VO methods
 - FE-HPM: filter estimation based on a semiblind deconvolution framework and HPM injection model [50]

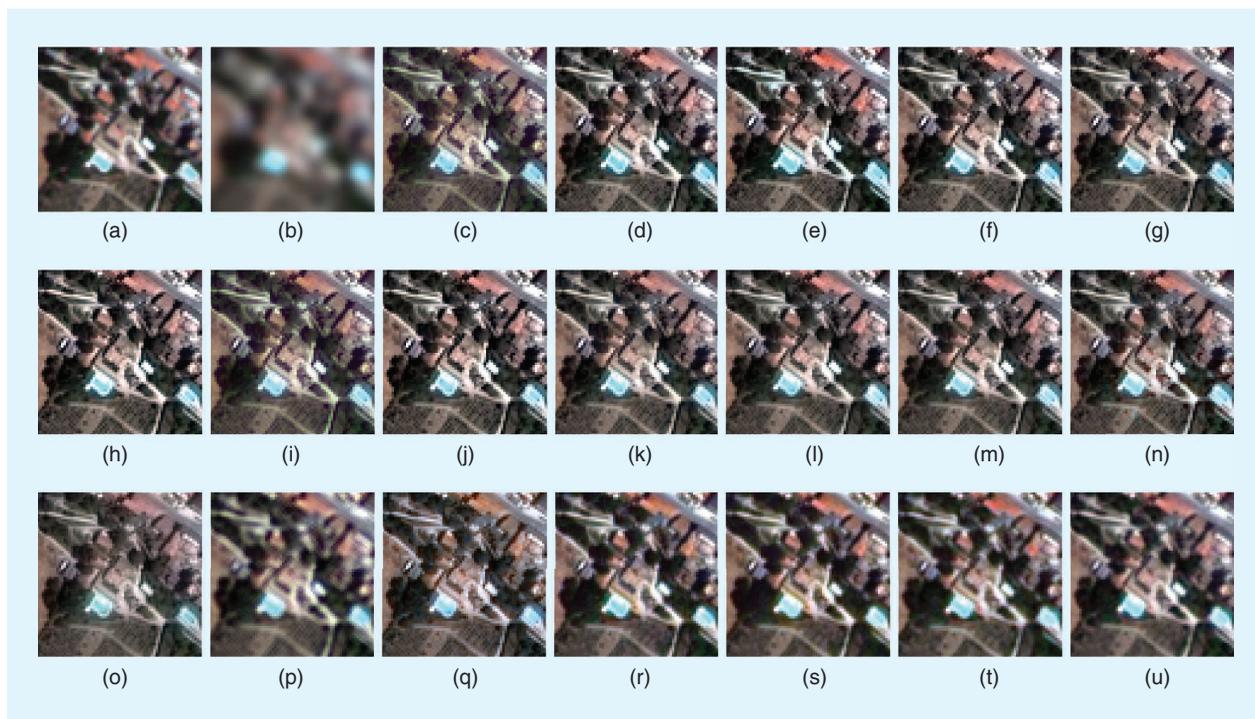


FIGURE 7. Close-ups of the main fusion results for the RR *GE-1* Collazzone (selected bands: red, green, and blue): (a) GT, (b) EXP, (c) BT-H, (d) BDS, (e) C-BDS, (f) BDS-PC, (g) C-GSA, (h) MTF-GLP-FS, (i) MTF-GLP-HPM-H, (j) MTF-GLP-HPM-R, (k) C-MTF-GLP-CBD, (l) MF, (m) FE-HPM, (n) SR-D, (o) PWMBF, (p) TV, (q) RR, (r) PNN, (s) PNN-IDX, (t) A-PNN, and (u) A-PNN-FT.

- SR-D: pansharpening based on sparse representation of injected details [88]
- PWMBF: model-based fusion using PCA and wavelets [79]
- TV: pansharpening based on TV [67]
- RR: model-based reduced-rank pansharpening [72], [134]
- ML methods
 - PNN: proposed in [91]
 - PNN-IDX: PNN augmenting the input by including several maps of nonlinear radiometric indexes as proposed in [91]
 - A-PNN: proposed in [95]
 - A-PNN-FT: proposed in [95].

A more detailed description of the methods can be found in the “Review of Pansharpening Methods” section and in the related references. For the sake of brevity, the setting of the parameters for each approach into the benchmark can be found by having a look at the default parameters in the related MATLAB toolbox distributed to the community.

REDUCED-RESOLUTION PERFORMANCE ASSESSMENT

This section focuses on the presentation of the experimental results at RR. The comparison will be performed class by class, splitting the section into four subsections, one for each class. For a wider (interclass) discussion, interested readers can refer to the “Discussion” section, which completes the discussion of the experimental results.

COMPONENT SUBSTITUTION

The products obtained through the CS methods have always been appreciated for the high quality of the spatial information, which is accurately transferred from the PAN image to the pansharpened image. The analysis conducted in this study confirms this consideration but highlights that a proper detail-extraction phase is required to obtain high performance. In fact, the GS method, which uses a simple average for constructing the low-pass PAN image from the MS channels, is not able to bear comparison with the other CS approaches.

A second crucial consideration regards the remarkable advances that have been obtained by the most recent implementations of the classical CS methods. In particular, partitioning the image in segments allows a more specific evaluation of the coefficients required by both the BDS and the GSA method. However, the context-adaptive BDS approach does not solve the difficulty of estimating the larger number of parameters involved in the fusion of the eight-band WV-3 images. On the contrary, the introduction of the constraints with the BDS-PC technique allows us to significantly increase the robustness of the method, avoiding physically unjustifiable solutions. Similarly, taking into account the characteristics of the image acquisition process permits us to achieve a high-quality fused product through the BT-H method, while the original BT version was classified as a low-performance approach in the previous study [14].

The visual analysis of the images, as in Figures 7 and 8, highlights that the improvements obtained by the recent

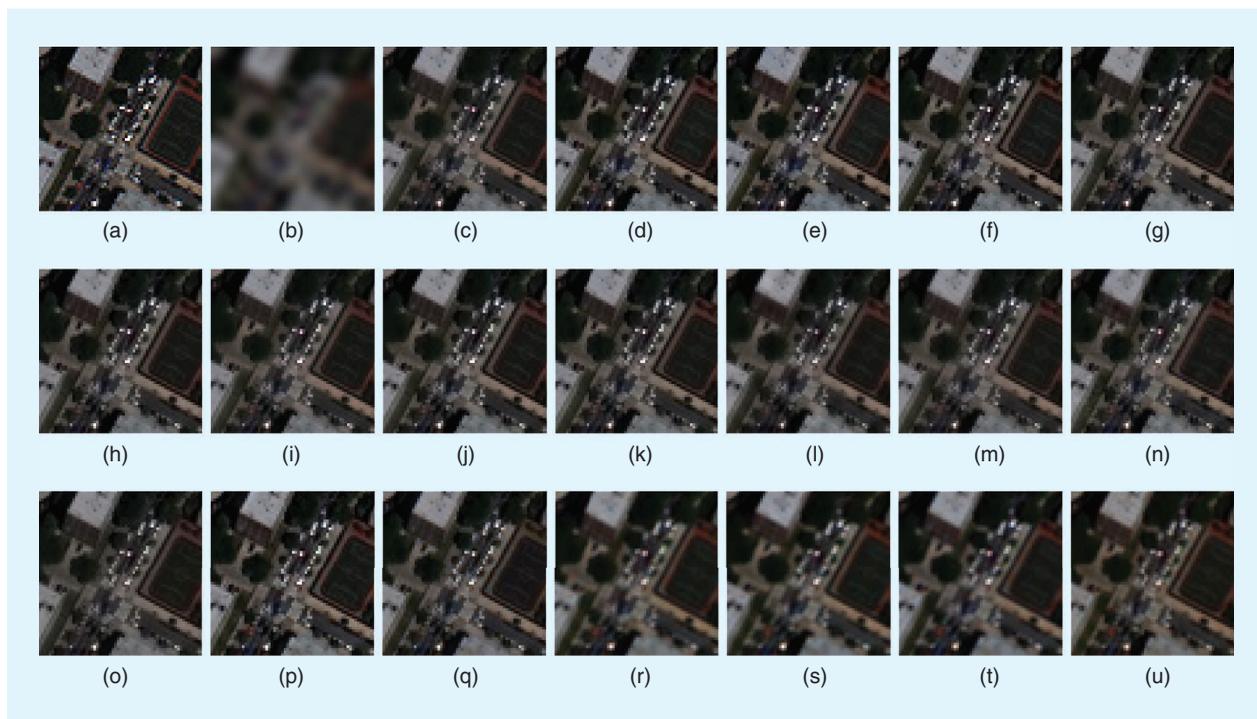


FIGURE 8. Close-ups of the main fusion results for the RR *WV-3* NY data set (selected bands: red, green, and blue): (a) GT, (b) EXP, (c) BT-H, (d) BDSD, (e) C-BDSD, (f) BDSD-PC, (g) C-GSA, (h) MTF-GLP-FS, (i) MTF-GLP-HPM-H, (j) MTF-GLP-HPM-R, (k) C-MTF-GLP-CBD, (l) MF, (m) FE-HPM, (n) SR-D, (o) PWMBF, (p) TV, (q) RR, (r) PNN, (s) PNN-IDX, (t) A-PNN, and (u) A-PNN-FT. (Satellite images courtesy of the Digital-Globe Foundation.)

approaches are related to a more targeted introduction of the PAN information into the original MS image, which is the result of taking into account the features of the acquisition sensors and the spatial characteristics of the illuminated scene. As a consequence, the details are more easily identifiable in the final products due to a higher contrast with respect to the surroundings. However, in some CS methods, this highlighting process can easily lead to overenhanced images, compromising the fidelity of the pansharpened product to the reference data.

MULTIRESOLUTION ANALYSIS

The key feature of the products obtained by applying MRA methods to aligned and aliasing-free images is the accurate preservation of the spectral characteristics of the original image, which is a visual counterpart of Wald's consistency property. In most cases, the results obtained using the recent algorithms based on a more accurate calculation of the injection coefficients are superior to those of the original implementations. However, as for MRA approaches, the most significant improvements are achieved by using a segmentation-based implementation, namely, through the C-MTF-GLP-CBD, and by modifying the MTF-GLP-HPM scheme by taking into consideration the haze contribution.

In accordance with the small difference among the quality indexes contained in Table 3, the visual differences among the images, as in Figures 7 and 8, are almost

negligible. The most significant quality enhancement regards the substantial reduction of the fusion artifacts introduced by the two cited C-MTF-GLP-CBD and MTF-GLP-HPM-H methods, which is particularly evident in the rural areas contained in the Collazzone data set.

VARIATIONAL OPTIMIZATION-BASED METHODS

By looking at experimental results in Table 3 obtained by VO techniques, it is possible to state that techniques relying on an acquisition model, such as TV and RR, provide results comparable with the best results obtained by the other pansharpening families. This can be explained by the effectiveness of the regularization (i.e., piecewise spatial smoothness for TV and low-rank representation for RR), which allows finding a suitable solution to the associated ill-posed inverse problem. The technique FE-HPM shows results that are comparable with MTF-based MRA techniques but not systematically better, which could be due to deviations of the sensor model considered with respect to the reality. The dictionary-based SR-D also reaches overall very good performance (always very close to the best VO approach for each test case), thanks to the image self-similarity, which is exploited in the sparse representation. The results obtained by PWMBF often stay below the other VO methods in terms of performance indexes for these tests. This can be associated with a slightly greater spectral and spatial distortion with respect to the other VO techniques.

TABLE 3. THE PERFORMANCE INDEXES (Q4/Q8, SAM MEASURED IN DEGREES, AND ERGAS) COMPUTED FOR THE THREE RR DATA SETS.

	<i>IK TOULOUSE</i>			<i>GE-1 COLLAZZONE</i>			<i>WV-3 NY</i>			
	<u>Q4</u>	<u>SAM [°]</u>	<u>ERGAS</u>	<u>Q4</u>	<u>SAM [°]</u>	<u>ERGAS</u>	<u>Q8</u>	<u>SAM [°]</u>	<u>ERGAS</u>	
GT	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	
EXP	0.4795	5.1823	6.3953	0.7850	2.3729	3.2092	0.6513	7.2118	8.1106	
CS	<u>BT-H</u>	0.9120	3.4491	2.9962	0.8865	2.5316	2.4135	0.9241	<u>6.4530</u>	3.9714
	<u>BDS</u>	0.9091	<u>2.9571</u>	<u>2.9307</u>	0.8850	2.6448	2.6387	0.9313	6.9330	3.9355
	<u>C-BDS</u>	0.9096	2.9595	2.9520	<u>0.8973</u>	2.5724	2.6169	0.9248	7.3269	4.3062
	<u>BDS-PC</u>	0.9094	2.9576	2.9309	0.8851	2.6369	2.6321	0.9327	6.8388	3.8905
	<u>GS</u>	0.7744	4.6757	4.6690	0.8227	2.6073	2.8448	0.8586	6.9743	5.2799
	<u>GSA</u>	0.9053	3.2310	3.0855	0.8703	3.0449	2.9043	0.9217	6.8037	4.0590
	<u>C-GSA</u>	0.9068	3.0756	3.0198	0.8852	<u>2.3943</u>	2.5085	0.9213	6.6967	4.0504
	<u>PRACS</u>	0.8845	3.8224	3.5279	0.8946	2.6709	<u>2.2256</u>	0.8972	7.2756	4.6838
	MRA	<u>AWLP</u>	0.9001	3.8647	3.2215	0.8752	2.6819	2.5292	0.9200	6.7638
<u>MTF-GLP</u>		0.8906	3.1934	3.3241	0.8660	2.5683	2.9581	0.9284	6.5710	<u>3.9013</u>
<u>MTF-GLP-FS</u>		0.9076	3.0906	3.0104	0.8754	2.9874	2.8254	0.9228	6.7650	4.0434
<u>MTF-GLP-HPM</u>		0.8964	3.1215	3.2068	0.8662	2.6734	2.9885	<u>0.9289</u>	6.6696	3.9032
<u>MTF-GLP-HPM-H</u>		<u>0.9121</u>	3.5111	3.0312	0.8874	2.5685	2.4208	0.9265	<u>6.4478</u>	3.9501
<u>MTF-GLP-HPM-R</u>		0.9105	3.1454	<u>2.9727</u>	0.8725	3.1349	2.9301	0.9228	7.0038	4.0692
<u>MTF-GLP-CBD</u>		0.9057	3.2102	3.0656	0.8735	3.0559	2.8790	0.9222	6.7780	4.0586
<u>C-MTF-GLP-CBD</u>		0.9039	<u>2.9279</u>	3.1068	0.8932	<u>2.3896</u>	<u>2.3836</u>	0.9211	6.6781	4.0627
<u>MF</u>		0.8766	3.6261	3.5924	<u>0.8946</u>	2.4802	2.4462	0.9019	6.7911	4.6377
VO	<u>FE-HPM</u>	0.8911	3.6422	3.3535	0.9003	2.4530	2.3438	0.9079	6.7763	4.3291
	<u>SR-D</u>	0.9108	2.9571	2.8708	0.9159	2.2700	2.1208	0.9113	6.6269	4.3472
	<u>PWMBF</u>	0.8756	3.2888	3.2093	0.8697	2.9468	2.4647	0.8972	7.4606	4.5811
	<u>TV</u>	0.9023	<u>2.8455</u>	2.9508	<u>0.9172</u>	2.4304	<u>1.9970</u>	<u>0.9277</u>	<u>6.6213</u>	4.0630
	<u>RR</u>	<u>0.9109</u>	2.9516	<u>2.8662</u>	0.8860	<u>2.2039</u>	2.2934	0.9256	6.8862	<u>4.0599</u>
ML	<u>PNN</u>	0.9132	2.8061	2.9078	0.9394	1.7729	1.7049	0.8937	6.3198	4.7539
	<u>PNN-IDX</u>	0.9067	3.0148	2.9808	0.9371	1.8382	1.7341	0.8972	<u>6.2281</u>	4.6880
	<u>A-PNN</u>	<u>0.9278</u>	<u>2.6846</u>	<u>2.8313</u>	0.9156	1.8923	2.1306	0.8946	6.2347	4.7359
	<u>A-PNN-FT</u>	0.9030	3.0385	3.1790	<u>0.9519</u>	<u>1.7697</u>	<u>1.5747</u>	<u>0.8991</u>	6.4752	<u>4.6136</u>

The best overall results are shown in boldface. The best results for each class of pansharpening algorithms are underlined.

MACHINE LEARNING

Let us start focusing on the two variants proposed in [91], PNN and PNN-IDX. Contrary to the experimental observations provided in [91], here PNN-IDX does not provide any performance gain over the baseline model PNN, giving even worse results on *IK* and *GE-1*. This can be justified by recalling that, in the original work, where PNN-IDX was recognized as the best model, both of these versions were trained on the same acquisitions used for tests (but separate cuts), whereas, in the current experimental frame train, test data are taken from different acquisitions. In [95], the use of indexes as additional input bands was also investigated employing different training and test acquisitions with results similar to ours. Therefore, we can safely conclude that the use of those additional features in input impacts the network robustness in real-world practical applications where the pretrained model has to be used as an off-the-shelf tool.

Moving to the advanced models proposed in [95], A-PNN and A-PNN-FT perform better than the PNN and PNN-IDX, on average. In particular, A-PNN achieves the

top performance on *IK*, not just among ML approaches, whereas A-PNN-FT is the best on *GE-1*. On *WV-3* images, instead, the various versions of PNN appear to lose their edge over non-ML methods. This is due in part to the fact that these ML models were originally conceived and optimized for *WV-2* images, while here we test their versions retrained on a *WV-3* data set of Rio de Janeiro.

At first glance, one may conclude that the FT is not as effective as expected. Indeed, this is only partly true. We may observe, in fact, that all models are pretrained by downgrading the resolution of the data sets assigned to the training phase. In addition to this, at inference time, A-PNN-FT performs additional training iterations on the RR version of the test image. As a consequence, if the test is performed on an image whose resolution has already been reduced (as in this evaluation framework), the downgrading applies twice, thus creating a resolution mismatch between the pretraining and the FT stages. This penalizes the FT version not only because of the poorer resolution of the training sample but also because of its reduced size, which can

TABLE 4. THE PERFORMANCE INDEXES (D_λ , D_s , AND HQNR) COMPUTED FOR THE THREE FR DATA SETS.

		<i>IK TOULOUSE FR</i>			<i>GE-1 COLLAZZONE FR</i>			<i>WV-3 NY FR</i>		
		D_λ	D_s	HQNR	D_λ	D_s	HQNR	D_λ	D_s	HQNR
	EXP	0.0563	0.1696	0.7836	0.0285	0.1156	0.8591	0.0562	0.1561	0.7964
CS	BT-H	0.0698	0.0847	0.8514	0.0387	0.0829	0.8816	0.0983	0.0829	0.8269
	BDS	0.0897	0.0642	0.8519	0.0408	0.1240	0.8402	0.1719	0.0484	0.7880
	C-BDS	0.0900	<u>0.0441</u>	0.8699	<u>0.0341</u>	0.0898	0.8792	0.1770	0.0598	0.7738
	BDS-PC	0.0883	0.0655	0.8520	0.0408	0.1254	0.8389	0.1554	<u>0.0251</u>	0.8234
	GS	0.1480	0.0581	0.8024	0.1688	<u>0.0605</u>	0.7809	0.1332	0.0500	0.8234
	GSA	0.0754	0.1053	0.8272	0.0562	0.1512	0.8011	0.1069	0.0793	0.8223
	C-GSA	0.0672	0.0808	0.8575	0.0501	0.1394	0.8175	0.1022	0.0747	0.8307
	PRACS	<u>0.0570</u>	0.0647	<u>0.8820</u>	0.0346	0.0754	<u>0.8926</u>	<u>0.0784</u>	0.0359	<u>0.8885</u>
MRA	AWLP	<u>0.0272</u>	0.0989	0.8765	<u>0.0194</u>	<u>0.0764</u>	<u>0.9057</u>	<u>0.0316</u>	0.0570	0.9132
	MTF-GLP	0.0293	0.0995	0.8742	0.0299	0.1504	0.8242	0.0342	0.0786	0.8899
	MTF-GLP-FS	0.0285	0.0930	0.8811	0.0270	0.1387	0.8381	0.0347	0.0740	0.8939
	MTF-GLP-HPM	0.0284	0.0928	0.8814	0.0318	0.1507	0.8223	0.0360	0.0738	0.8929
	MTF-GLP-HPM-H	0.0286	0.0761	0.8975	0.0196	0.0779	0.9040	0.0344	0.0831	0.8854
	MTF-GLP-HPM-R	0.0277	0.0872	0.8874	0.0283	0.1384	0.8372	0.0356	0.0679	0.8989
	MTF-GLP-CBD	0.0285	0.0931	0.8810	0.0275	0.1397	0.8367	0.0349	0.0722	0.8955
	C-MTF-GLP-CBD	0.0356	<u>0.0317</u>	<u>0.9338</u>	0.0235	0.1243	0.8551	0.0375	<u>0.0407</u>	<u>0.9233</u>
	MF	0.0423	0.0784	0.8826	0.0387	0.1434	0.8234	0.0381	0.0617	0.9026
	VO	FE-HPM	0.0289	0.0816	0.8919	0.0305	0.1471	0.8268	0.0347	0.0693
SR-D		0.0173	0.0444	<u>0.9391</u>	0.0082	0.0248	0.9672	0.0199	0.0369	0.9440
PWMBF		0.0844	0.1026	0.8217	0.0691	0.0480	0.8862	0.0716	0.1082	0.8279
TV		0.0471	<u>0.0303</u>	0.9240	0.0553	0.0534	0.8943	0.0234	<u>0.0252</u>	0.9520
ML	RR	0.0803	0.4136	0.5393	0.0680	0.0633	0.8731	0.0956	0.0597	0.8504
	PNN	0.0375	0.0539	0.9107	0.0225	0.0767	0.9026	0.0535	0.0639	0.8860
	PNN-IDX	0.0609	0.0334	0.9077	0.0253	0.0581	0.9180	0.0599	0.1037	0.8426
	A-PNN	<u>0.0342</u>	0.0399	0.9273	<u>0.0143</u>	0.0568	0.9298	<u>0.0463</u>	0.0756	0.8816
	A-PNN-FT	0.0344	0.0265	0.9399	0.0146	<u>0.0451</u>	<u>0.9410</u>	0.0523	0.0188	<u>0.9299</u>

The best overall results are shown in boldface. The best results for each class of pansharpening algorithms are underlined.

become very small if the test image is not sufficiently large. In our case, the FT eventually operates on a 128×128 (PAN scale) input. Of course, what is really important is that this mismatch does not occur in practical applications at FR.

FULL-RESOLUTION PERFORMANCE ASSESSMENT

This section focuses on the experimental results at FR. The comparison will be performed by splitting the section into four subsections, one for each class. For a broader (inter-class) discussion, interested readers can refer to the "Discussion" section, which completes the discussion of the experimental results.

COMPONENT SUBSTITUTION

The analyses of the FR quality indexes in Table 4 are in substantial accordance with the results obtained at RR. However, the presence of details at a higher resolution gives evidence of different characteristics. In particular, it becomes clear that the improvements caused by the segmentation algorithms are mainly due to a significant reduction of the spatial distortion. The most evident difference from the RR protocol is in regards to the PRACS method, which obtains

the best results in terms of the chosen quality indexes without reference, rewarding the optimal balance between detail enhancement and consistency with respect to the original data.

In fact, as is also clear from Figure 9, most algorithms perform strong enhancement of the local contrast entailing the presence of zones with unnatural colors (see, for example, the colors of the roofs in the C-BDS pansharpened product related to the Toulouse data set).

MULTIRESOLUTION ANALYSIS

The performance of the algorithms at FR strictly matches those obtained at RR, underlining the significant improvement achievable by running the MTF-GLP-CBD on image segments and by correcting the HPM scheme to take into account the atmospheric effects.

Thanks to the presence of vast rural and vegetated areas, the Collazzone data set highlights the importance of considering the path-radiance contribution. Figure 10(c), related to the MTF-GLP-HPM-H method, shows the significant noise reduction obtained by excluding the haze from the pansharpening process. The corresponding quality

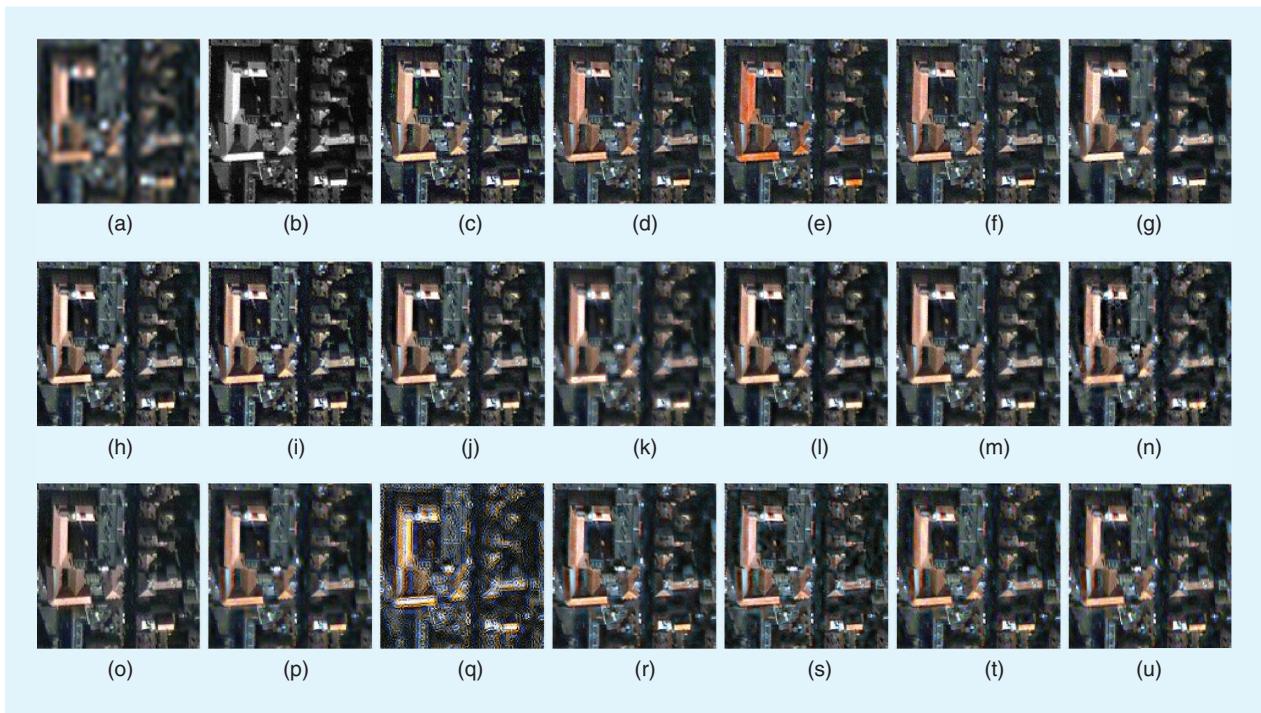


FIGURE 9. Close-ups of the main fusion results for the FR *IK* Toulouse (selected bands: red, green, and blue): (a) EXP, (b) PAN, (c) BT-H, (d) BDDSD, (e) C-BDDSD, (f) BDDSD-PC, (g) C-GSA, (h) MTF-GLP-FS, (i) MTF-GLP-HPM-H, (j) MTF-GLP-HPM-R, (k) C-MTF-GLP-CBD, (l) MF, (m) FE-HPM, (n) SR-D, (o) PWMBF, (p) TV, (q) RR, (r) PNN, (s) PNN-IDX, (t) A-PNN, and (u) A-PNN-FT.

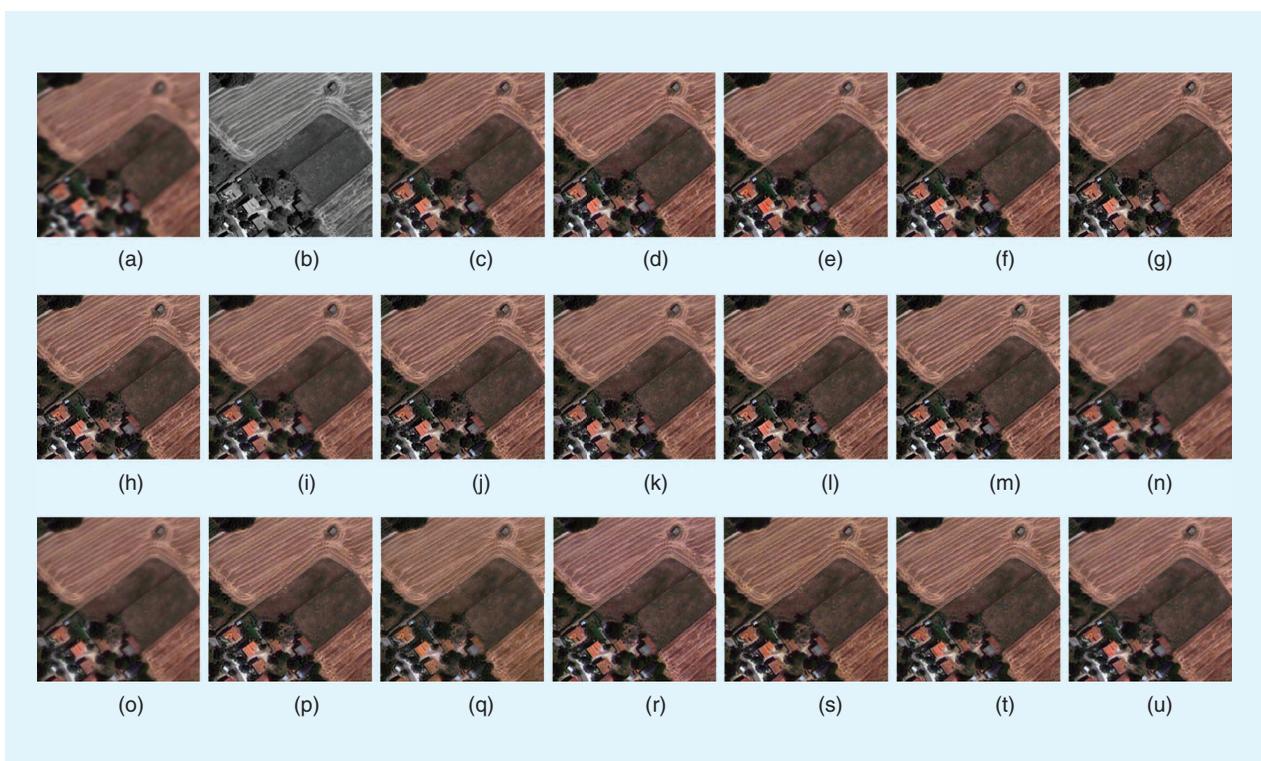


FIGURE 10. Close-ups of the main fusion results for the FR *GE-1* Collazzone (selected bands: red, green, and blue): (a) EXP, (b) PAN, (c) BT-H, (d) BDDSD, (e) C-BDDSD, (f) BDDSD-PC, (g) C-GSA, (h) MTF-GLP-FS, (i) MTF-GLP-HPM-H, (j) MTF-GLP-HPM-R, (k) C-MTF-GLP-CBD, (l) MF, (m) FE-HPM, (n) SR-D, (o) PWMBF, (p) TV, (q) RR, (r) PNN, (s) PNN-IDX, (t) A-PNN, and (u) A-PNN-FT.

TABLE 5. THE COMPUTATIONAL TIMES (IN SECONDS) OF THE COMPARED ALGORITHMS AS A FUNCTION OF THE FUSED IMAGE SIZE.

		FOUR-BAND MS				EIGHT-BAND MS			
		256	512	1024	2048	256	512	1024	2048
	EXP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CS	BT-H	0.0	0.1	0.3	1.2	0.0	0.1	0.4	1.5
	BDS	0.0	0.1	0.3	1.2	0.1	0.1	0.5	2.0
	C-BDS	0.3	1.4	6.4	26.9	0.6	2.6	10.7	48.2
	BDS-PC	0.0	0.1	0.3	1.2	0.1	0.2	0.6	2.5
	GS	0.0	0.1	0.3	1.3	0.0	0.2	0.6	2.4
	GSA	0.0	0.2	0.7	3.0	0.1	0.3	1.1	4.3
	C-GSA	0.3	1.6	7.0	36.6	0.3	1.5	8.0	31.3
	PRACS	0.1	0.3	1.2	4.9	0.2	0.9	3.9	15.4
MRA	AWLP	0.1	0.4	1.8	7.5	0.1	0.8	3.6	14.7
	MTF-GLP	0.1	0.3	1.0	4.5	0.2	0.5	1.9	8.6
	MTF-GLP-FS	0.1	0.2	0.7	2.9	0.1	0.4	1.3	5.7
	MTF-GLP-HPM	0.1	0.3	1.0	4.4	0.2	0.5	2.0	8.5
	MTF-GLP-HPM-H	0.1	0.2	0.9	3.8	0.1	0.4	1.5	6.5
	MTF-GLP-HPM-R	0.1	0.2	0.6	2.7	0.1	0.3	1.2	5.3
	MTF-GLP-CBD	0.1	0.2	0.7	3.1	0.1	0.4	1.4	6.1
	C-MTF-GLP-CBD	0.3	1.4	7.4	30.3	0.3	1.8	8.6	33.3
VO	MF	0.0	0.1	0.6	2.2	0.1	0.3	1.1	4.5
	FE-HPM	0.2	0.6	2.6	10.8	0.2	1.0	3.9	16.0
	SR-D	1.2	5.1	40.8	553.4	2.1	9.4	92.8	992.0
	PWMBF	0.2	0.9	3.9	17.4	0.3	1.0	4.0	18.2
	TV	0.4	2.7	11.6	46.9	4.4	26.7	116.9	472.8
ML	RR	28.0	207.9	1053.3	4482.8	79.8	326.5	1309.0	5678.7
	PNN	0.5	0.7	1.6	5.0	0.6	0.9	2.4	8.1
	PNN-IDX	0.5	0.7	1.7	5.5	0.6	1.0	2.8	9.7
	A-PNN	0.6	0.8	1.7	5.2	0.6	1.0	2.6	8.6
	A-PNN-FT	4.2	6.5	16.3	55.5	4.6	7.9	21.6	77.9
MLGPU	PNN	0.5	0.5	0.7	1.7	0.5	0.6	1.0	2.7
	PNN-IDX	0.4	0.5	0.8	2.2	0.5	0.6	1.2	3.7
	A-PNN	0.5	0.6	0.8	1.9	0.5	0.6	1.1	3.1
	A-PNN-FT	3.8	3.9	4.6	7.6	4.0	4.2	5.4	10.4

indexes in Table 4 confirm the best results obtained by this method.

VARIATIONAL OPTIMIZATION-BASED METHODS

The results reported in Table 4 for the VO techniques overall consolidate the outcomes of the analysis at RR. Overall good results can be pointed out by both TV and SR-D. In particular, the TV is the best method for the WV-3 NY data set thanks to its spatial regularization, which can positively match the urban and human-made structures. SR-D also obtains high-performance scores, overall outperforming all other VO techniques for the Toulouse and the GE-1 data sets. The results of the FE-HPM and PWMBF techniques are, in general, behind those algorithms, as also shown at RR. Finally, the RR method often shows the opposite behavior with respect to the RR assessment (e.g., it was the best VO method for IK Toulouse at RR but got the lowest performance for the related FR data set).

MACHINE LEARNING

The overall behavior of the ML methods at FR is rather satisfying. If we focus on the HQNR figure, which summarizes spatial and spectral distortions, ML scores fall between 0.9 and 0.97 for the Toulouse and Collazzone data sets, with A-PNN-FT always above 0.93. Lower performance can be remarked only on the NY data sets, following the results obtained at RR. In more detail, on IK and GE-1 data, a progressive performance gain is observed when moving from the basic model PNN to the most advanced one, A-PNN-FT. A bit less obvious is the analysis for WV-3. PNN, PNN-IDX, and A-PNN, in fact, show a performance drop. As already remarked, PNN-IDX is more sensitive to the (statistical) alignment between the training and test data sets than PNN. As a consequence, its performance drop on WV-3 indicates that such a statistical shift is larger in this case. This intuition is confirmed by the large HQNR gain of A-PNN-FT over A-PNN (recall that the former is a refinement of the latter), about 5%, whereas it is

smaller than 2% on the other two data sets. Of course, the pretty good performance of A-PNN-FT is achieved thanks to a supplemental training phase that requires additional computational efforts.

COMPUTATIONAL ANALYSIS

A crucial point of this study is the analysis of the computational effort required by the compared algorithms. While classical methods (CS and MRA) take approximately the same time to generate the final product, third-generation (VO and ML) approaches could be characterized by a longer execution time, which is quite algorithm dependent.

We report in Table 5 the execution times of the compared methods, calculated using a workstation equipped with an Intel Core I7 3.2-GHz processor and a Titan XP GPU. A general trend is the linear increase of the computational time with the number of pixels (namely, with the square of the image size indicated in the table).

Taking into consideration the values obtained by the CS and MRA algorithms, we can note that most recent implementations require a computational effort very similar to the original versions. The unique exception is represented by the methods based on image segmentation, which are characterized by higher computational times. The increase is due to the preliminary classification phase and to the additional work required by the segment-by-segment fusion. While the first issue can be limited when a fusion algorithm is repetitively used to fuse data acquired over the same area, the second problem always introduces an additional effort, which grows with the number of segments.

The key considerations involve the VO methods, which are based on an optimization phase. Some of the compared approaches (FE-HPM and PWMBF) are more efficient, since they are based on credited models for data fusion whose parameters can be estimated through efficient procedures. The remaining three VO techniques are the most time-consuming ones. The SR-D method, which includes a demanding phase of dictionary construction and a patch-based optimization of the representation coefficients, is characterized by a computational time growth (a function of the number of pixels) that exceeds the linear trend shown by most of the compared algorithms. However, this effect can be reduced by partitioning the data into subimages. This gives back to the linear trend without implying a significant reduction of the performance.

Finally, we report in Table 5 the computational times related to the ML techniques. In particular, we first indicate the effort required by the algorithms executed using the workstation CPU, which can be straightforwardly compared to those of the other algorithms. The values related to the approaches without FT are quite similar to classical methods. On the contrary, the A-PNN-FT takes longer than CS and MRA approaches to produce the pansharpened product. However, it is remarkable that the FT included in the A-PNN-FT approach does not turn out to be an intractable step also in the case of large-sized images. As a

TABLE 6. THE TRAINING AND VALIDATION SAMPLES.

SENSOR	SITE	TRAINING	VALIDATION
<i>IK</i>	Caserta	14,400 × (33 × 33)	7,168 × (33 × 33)
<i>GE-1</i>	Caserta	14,400 × (33 × 33)	7,168 × (33 × 33)
<i>WV-3</i>	Rio de Janeiro	10,800 × (33 × 33)	3,600 × (33 × 33)

matter of fact, the most natural way to carry out the ML approaches is by means of a GPU, which allows realizing a parallel implementation of the network. The values in the last four rows of Table 5 confirm that a significant reduction of the execution times can be achieved by employing the workstation GPU.

The number of training samples for a given CNN model must be sufficiently large to prevent overfitting and ensure generality. On the other hand, a too-large data set would require excessive training time. Aware of this tradeoff, the authors of the several PNN versions [91], [95] have collected training and validation samples as described in Table 6. For each sensor, a single, sufficiently large image was selected. Once rescaled according to Wald's protocol, tens of thousands of small (33 × 33) patches were randomly cropped from them to create disjointed sets for training and validation. Moreover, the training sets were organized in minibatches of 128 elements for an efficient implementation of the stochastic gradient descent algorithm used for optimization. All networks were trained using a fixed number of iterations (10⁶) over the minibatches, corresponding to about ten thousand epochs (passages over the whole training data set), and took approximately the same computational time of about ten hours using a GTX Titan X 12-GB GPU card.

DISCUSSION

A benchmark consisting of 26 pansharpening approaches has been proposed and tested on challenging scenarios, from urban to rural areas, considering many commonly used commercial sensors. Generally speaking, the experimental results show very high performance for almost all of the methods in the benchmark. In fact, except for the outdated GS method (included for aiding possible meta-analysis), all of the techniques got values of the overall quality index, $Q2^n$, greater than 0.87; see Table 3. A general trend over the RR data sets can also be drawn. Indeed, better results (with many algorithms getting $Q2^n$ values around 0.92) can be remarked when *WV-3* data are fused. Instead, lower performance (with many $Q2^n$ values lower than 0.90) is obtained for the *GE-1* Collazzone data set. These results can be justified considering that 1) many researchers rely upon the assessment of data capturing urban scenarios and involving data acquired by the WorldView-2/-3 satellites and 2) the performance assessment on rural scenarios (as the *GE-1* Collazzone data set) is often neglected in the literature. Furthermore, this trend at RR is confirmed at FR.

After a general discussion about the results obtained by the proposed benchmark, the following remarks are related to the interclass comparison. By focusing on the results at RR and, in particular, by analyzing the overall quality index, $Q2^n$, the A-PNN method gets the best results on the *IK* Toulouse data set, followed, in order, by BT-H, MTF-GLP-HPM-H, and RR. Thus, comparable performance can be obtained in this case among the best methods in the four classes. Instead, the gap between classical methods (i.e., CS and MRA) and the third-generation approaches (i.e., VO and ML) increases when rural scenarios (e.g., the *GE-1* Collazzone data set) are considered. In that case, ML approaches clearly represent the best solution (in particular, the A-PNN-FT is the best method), followed by some VO approaches (i.e., TV and SR-D). These results encourage future research in the direction of fusing rural scenarios using CS or MRA approaches. The last test case is instead related to the fusion of the eight-band WV-3 NY data set. Table 3 shows an opposite trend with respect to the *GE-1* Collazzone data set. In fact, the best approach is represented by the BDSD-PC, followed by several other CS and MRA methods. Comparable performance out of the CS and the MRA categories is only obtained by the RR and the TV methods, both belonging to the VO class. Instead, ML methods get lower performance, with $Q2^n$ indexes lower than 0.9. This is due to the fact that the related models were originally conceived and optimized for WV-2 images, while we test their versions retrained on a sole WV-3 data set. The last remark about the outcomes at RR is related to the comparison between CS and MRA. Indeed, the optimized versions of these methods in both categories converge toward the same (high) performance in all of the test cases.

Further analysis is performed by having a look at the FR outcomes that generally corroborate the ones at RR. This is particularly true for the intraclass comparison. Instead, when methods belonging to different fusion philosophies are compared, the FR HQNR index shows some limitations because of the absence of a GT image for an accurate performance assessment, thus often experimenting with inaccurate values for the spatial distortion index, D_s . For instance, in the case of the WV-3 NY data set, MRA methods got better results than CS methods (instead, at RR, a CS approach was the best choice), and the CS and MRA methods obtained lower performance than A-PNN-FT and some VO algorithms (it is easy to see the opposite behavior at RR). A final comment about the FR outcomes reported in Table 4 is that the A-PNN-FT method shows its superiority, as measured by the overall quality index HQNR, with respect to the other approaches, on the *IK* Toulouse data set. Instead, the advantages are reduced when the FR WV-3 NY data set is considered. In this case, the best performance is obtained by the TV method, followed, in order, by SR-D and A-PNN-FT. Generally speaking, the A-PNN-FT and some VO approaches, such as TV and SR-D, are favorites at FR, showing remarkably higher HQNR values than CS methods and usually better performance than MRA methods.

Finally, computational analysis is performed in the "Computational Analysis" section. For this study, CS and MRA approaches represent a sort of baseline, not requiring a high computational burden (except for segmentation-based approaches). Moreover, they do not usually need any parameter tuning (as required by VO methods); nor do they require a training phase (as required by ML methods). Generally speaking, we can say that some VO approaches are quite time consuming (e.g., the RR method). It is useful to remark that VO methods are very sensitive to the choice of hyperparameters, both in terms of computational time and performance. Accordingly, they often require an accurate optimization phase that might hide their true potential. For this reason, VO methods could become practical if the problem of selecting good hyperparameters is quickly solved.

A reduced effort with respect to VO methods is instead required by ML approaches, leveraging the use of a GPU architecture. The unique exception (requiring more processing time than the other ML-based methods) is given by the A-PNN-FT that exploits an FT phase to improve the performance. The last note is about the methods of the ML class. Indeed, in this case, the time-consuming phase is the training of the underlying models for each fusion problem to be addressed. Indeed, it requires special hardware (GPUs) to be used, involving many samples to properly train those models to reach high performance in the operational phase.

SUMMARY

This article investigated several state-of-the-art pansharp-ening algorithms belonging to different classes. In particular, a critical comparison, including both classical (i.e., CS and MRA) and the third-generation (i.e., VO and ML) approaches, has been presented. To our knowledge, this is the first attempt to fairly compare all of these heterogeneous methods using a common framework for the assessment. Data acquired by three commercial sensors (i.e., *IK*, *GE-1*, and WV-3) over different scenarios, from urban to rural areas, have been exploited. Both the assessments at RR and at FR have been considered. Furthermore, a computational analysis has been provided. Very high performance (both at FR and at RR) has been shown by some ML approaches. However, this requires a huge number of data samples and a time-demanding training phase for the underlying models. Furthermore, some methods in the VO class, such as TV and SR-D, obtained state-of-the-art performance but with a higher computational burden than classical (CS and MRA) approaches. Finally, high performance has been shown by many recent advanced CS and MRA methods, in particular, on WV-3 data, where BDSD-PC turns out to be the best approach at RR. However, this trend is not confirmed when rural scenarios are involved (see the results on the *GE-1* Collazzone data set), suggesting an additional effort in developing new CS and MRA techniques that are also able to deal with this (less explored) problem. A concluding remark is about

the assessment at FR. In fact, the results often corroborate the ones at RR. However, when an interclass comparison involving approaches referring to very different fusion philosophies is performed, some inconsistencies can be pointed out, pushing future research toward this debated scientific direction.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the NVIDIA Corporation with the donation of the Titan XP GPU used for this research and the DigitalGlobe Foundation for providing the *WorldView-3* NY data set.

AUTHOR INFORMATION

Gemine Vivone (gemine.vivone@imaa.cnr.it, gvivone@unisa.it) is with the Institute of Methodologies for Environmental Analysis, Tito Scalo, Italy, and the Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno, Fisciano, Italy. He is a Senior Member of IEEE.

Mauro Dalla Mura (mauro.dalla-mura@gipsa-lab.grenoble-inp.fr) is with GIPSA-Lab, Grenoble Institute of Technology, Saint Martin d'Hères, France, and with the Tokyo Tech World Research Hub Initiative, School of Computing, Tokyo Institute of Technology, Japan. He is a Senior Member of IEEE.

Andrea Garzelli (andrea.garzelli@unisi.it) is with the Department of Information Engineering and Mathematics, University of Siena, Italy. He is a Senior Member of IEEE.

Rocco Restaino (restaino@unisa.it) is with the Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno, Fisciano, Italy. He is a Senior Member of IEEE.

Giuseppe Scarpa (giscarpa@unina.it) is with the Department of Electrical Engineering and Information Technology, University Federico II, Naples, Italy. He is a Senior Member of IEEE.

Magnus O. Ulfarsson (mou@hi.is) is with the Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik. He is a Senior Member of IEEE.

Luciano Alparone (luciano.alparone@unifi.it) is with the Department of Information Engineering, University of Florence, Italy.

Jocelyn Chanussot (jocelyn.chanussot@grenoble-inp.fr) is with the University of Grenoble Alpes, France, and the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. He is a Fellow of IEEE.

REFERENCES

- [1] M. Schmitt and X. X. Zhu, "Data fusion and remote sensing: An ever-growing relationship," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 4, pp. 6–23, Dec. 2016. doi: 10.1109/MGRS.2016.2561021.
- [2] J. K. Gilbertson, J. Kemp, and A. Van Niekerk, "Effect of pansharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques," *Comput. Electron. Agric.*, vol. 134, pp. 151–159, Mar. 2017. doi: 10.1016/j.compag.2016.12.006.
- [3] P. Du, S. Liu, J. Xia, and Y. Zhao, "Information fusion techniques for change detection from multi-temporal remote sensing images," *Inform. Fusion*, vol. 14, no. 1, pp. 19–27, Jan. 2013. doi: 10.1016/j.inffus.2012.05.003.
- [4] Y. Qu, H. Qi, B. Ayhan, C. Kwan, and R. Kidd, "DOES multispectral/hyperspectral pansharpening improve the performance of anomaly detection?" in *Proc. 2017 IEEE Int. Geoscience Remote Sensing Symp. (IGARSS)*, pp. 6130–6133. doi: 10.1109/IGARSS.2017.8128408.
- [5] G. Vivone, R. Restaino, G. Licciardi, M. Dalla Mura, and J. Chanussot, "Multiresolution analysis and component substitution techniques for hyperspectral pansharpening," in *Proc. 2014 IEEE Geoscience Remote Sensing Symp.*, pp. 2649–2652. doi: 10.1109/IGARSS.2014.6947018.
- [6] L. Loncan et al., "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sept. 2015. doi: 10.1109/MGRS.2015.2440094.
- [7] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hyper-sharpening: A first approach on SIM-GA data," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 6, pp. 3008–3024, June 2015. doi: 10.1109/JSTARS.2015.2440092.
- [8] D. Picone, R. Restaino, G. Vivone, P. Addesso, M. Dalla Mura, and J. Chanussot, "Band assignment approaches for hyperspectral sharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 739–743, May 2017. doi: 10.1109/LGRS.2017.2677087.
- [9] J. Marcello, E. Ibarrola-Ulzurrun, C. Gonzalo-Martín, J. Chanussot, and G. Vivone, "Assessment of hyperspectral sharpening methods for the monitoring of natural areas using multiplatform remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8208–8222, 2019. doi: 10.1109/TGRS.2019.2918932.
- [10] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, May 2017. doi: 10.1109/LGRS.2017.2668299.
- [11] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli, *Remote Sensing Image Fusion*. Boca Raton, FL: CRC Press, 2015.
- [12] P. Ghamisi et al., "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 6–39, Mar. 2019. doi: 10.1109/MGRS.2018.2890023.
- [13] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Twenty-five years of pansharpening: A critical review and new developments," in *Signal and Image Processing for Remote Sensing*, C.-H. Chen, Ed., 2nd ed. Boca Raton, FL: CRC Press, 2012, pp. 533–548.
- [14] G. Vivone et al., "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015. doi: 10.1109/TGRS.2014.2361734.
- [15] H. Ghassemian, "A review of remote sensing image fusion methods," *Inform. Fusion*, vol. 32, pp. 75–89, Nov. 2016. doi: 10.1016/j.inffus.2016.03.003.
- [16] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inform. Fusion*, vol. 33, pp. 100–112, Jan. 2017. doi: 10.1016/j.inffus.2016.05.004.

- [17] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu, "Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges," *Inform. Fusion*, vol. 46, pp. 102–113, Mar. 2019. doi: 10.1016/j.inffus.2018.05.006.
- [18] A. Garzelli, "A review of image fusion algorithms based on the super-resolution paradigm," *Remote Sens.*, vol. 8, no. 10, pp. 797:1–797:20, Sept. 2016. doi: 10.3390/rs8100797.
- [19] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017. doi: 10.1109/MGRS.2017.2762307.
- [20] W. Carper, T. Lillesand, and R. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, Apr. 1990.
- [21] P. S. Chavez Jr., S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, Mar. 1991.
- [22] P. S. Chavez Jr. and A. W. Kwarteng, "Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 3, pp. 339–348, Mar. 1989.
- [23] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, May 1992.
- [24] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6 011 875, 2000.
- [25] X. Kang, S. Li, and J. A. Benediktsson, "Pansharpening with matting model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5088–5099, Aug. 2014. doi: 10.1109/TGRS.2013.2286827.
- [26] M. Ghahremani and H. Ghassemian, "Nonlinear IHS: A promising method for pan-sharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1606–1610, Nov. 2016. doi: 10.1109/LGRS.2016.2597271.
- [27] G. A. Licciardi, M. M. Khan, J. Chanussot, A. Montanvert, L. Condat, and C. Jutten, "Fusion of hyperspectral and panchromatic images using multiresolution analysis and nonlinear PCA band reduction," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 207:1–207:17, Sept. 2012. doi: 10.1186/1687-6180-2012-207.
- [28] Q. Xu, B. Li, Y. Zhang, and L. Ding, "High-fidelity component substitution pansharpening by the fitting of substitution data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7380–7392, Nov. 2014.
- [29] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, Jan. 2011. doi: 10.1109/TGRS.2010.2051674.
- [30] B. Aiuzzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007. doi: 10.1109/TGRS.2007.901007.
- [31] L. Alparone, A. Garzelli, and G. Vivone, "Inter-sensor statistical matching for pansharpening: Theoretical issues and practical solutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4682–4695, Aug. 2017. doi: 10.1109/TGRS.2017.2697943.
- [32] R. Restaino, M. Dalla Mura, G. Vivone, and J. Chanussot, "Context-adaptive pansharpening based on image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 753–766, Feb. 2017. doi: 10.1109/TGRS.2016.2614367.
- [33] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 228–236, Jan. 2008. doi: 10.1109/TGRS.2007.907604.
- [34] A. Garzelli, "Pansharpening of multispectral images based on nonlocal parameter optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2096–2107, Apr. 2015. doi: 10.1109/TGRS.2014.2354471.
- [35] G. Vivone, "Robust band-dependent spatial-detail approaches for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6421–6433, Sept. 2019. doi: 10.1109/TGRS.2019.2906073.
- [36] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Commun. Lett.*, vol. 31, no. 4, pp. 532–540, Apr. 1983. doi: 10.1109/TCOM.1983.1095851.
- [37] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, July 1989. doi: 10.1109/34.192463.
- [38] M. J. Shensa, "The discrete wavelet transform: Wedding the à trous and Mallat algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992. doi: 10.1109/78.157290.
- [39] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*, vol. 103, A. Antoniadis and G. Oppenheim, Eds. New York: Springer-Verlag, 1995, pp. 281–299.
- [40] J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, June 2002. doi: 10.1109/TIP.2002.1014998.
- [41] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005. doi: 10.1109/TIP.2005.859376.
- [42] J. G. Liu, "Smoothing filter based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, Dec. 2000. doi: 10.1080/014311600750037499.
- [43] B. Aiuzzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002. doi: 10.1109/TGRS.2002.803623.
- [44] J. Núñez, X. Otazu, O. Fors, A. Prades, V. Palà, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999. doi: 10.1109/36.763274.
- [45] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inform.*

- Fusion*, vol. 8, no. 2, pp. 143–156, Sept. 2007. doi: 10.1016/j.inffus.2006.02.001.
- [46] V. P. Shah, N. H. Younan, and R. L. King, “An efficient pan-sharpening method via a combined adaptive-PCA approach and contourlets,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008. doi: 10.1109/TGRS.2008.916211.
- [47] B. Aiuzzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, “MTF-tailored multiscale fusion of high-resolution MS and Pan imagery,” *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006. doi: 10.14358/PERS.72.5.591.
- [48] L. Alparone, S. Baronti, B. Aiuzzi, and A. Garzelli, “Spatial methods for multispectral pansharpening: Multiresolution analysis demystified,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2563–2576, May 2016. doi: 10.1109/TGRS.2015.2503045.
- [49] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, “MTF-based deblurring using a Wiener filter for CS and MRA pansharpening methods,” *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 9, no. 6, pp. 2255–2269, June 2016. doi: 10.1109/JSTARS.2016.2546061.
- [50] G. Vivone et al., “Pansharpening based on semiblind deconvolution,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1997–2010, Apr. 2015. doi: 10.1109/TGRS.2014.2351754.
- [51] H. R. Shahdoosti and H. Ghassemian, “Fusion of MS and PAN images preserving spectral quality,” *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 611–615, 2014. doi: 10.1109/LGRS.2014.2353135.
- [52] G. Vivone, P. Addesso, R. Restaino, M. Dalla Mura, and J. Chanussot, “Pansharpening based on deconvolution for multiband filter estimation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 540–553, Jan. 2019. doi: 10.1109/TGRS.2018.2858288.
- [53] G. Vivone and J. Chanussot, “Fusion of short-wave infrared and visible near-infrared WorldView-3 data,” *Inform. Fusion*, vol. 61, pp. 71–83, Sept. 2020. doi: 10.1016/j.inffus.2020.03.012.
- [54] S. Zheng, W. Shi, J. Liu, and J. Tian, “Remote sensing image fusion using multiscale mapped LS-SVM,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1313–1322, May 2008. doi: 10.1109/TGRS.2007.912737.
- [55] R. Restaino, G. Vivone, M. Dalla Mura, and J. Chanussot, “Fusion of multispectral and panchromatic images based on morphological operators,” *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2882–2895, June 2016. doi: 10.1109/TIP.2016.2556944.
- [56] G. Vivone, L. Alparone, A. Garzelli, and S. Lolli, “Fast reproducible pansharpening based on instrument and acquisition modeling: AWLP revisited,” *Remote Sens.*, vol. 11, no. 19, pp. 2315:1–2315:23, Oct. 2019. doi: 10.3390/rs11192315.
- [57] R. Restaino, G. Vivone, P. Addesso, and J. Chanussot, “A pansharpening approach based on multiple linear regression estimation of injection coefficients,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 102–106, Jan. 2020. doi: 10.1109/LGRS.2019.2914093.
- [58] Y. Zhang, “A new merging method and its spectral and spatial effects,” *Int. J. Remote Sens.*, vol. 20, no. 10, pp. 2003–2014, 1999. doi: 10.1080/014311699212317.
- [59] S. Lolli, L. Alparone, A. Garzelli, and G. Vivone, “Haze correction for contrast-based multispectral pansharpening,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2255–2259, Dec. 2017. doi: 10.1109/LGRS.2017.2761021.
- [60] A. Garzelli, B. Aiuzzi, L. Alparone, S. Lolli, and G. Vivone, “Multispectral pansharpening with radiative transfer-based detail-injection modeling for preserving changes in vegetation cover,” *Remote Sens.*, vol. 10, no. 8, pp. 1308:1–1308:18, Aug. 2018. doi: 10.3390/rs10081308.
- [61] X. Otazu, M. González-Audicana, O. Fors, and J. Núñez, “Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005. doi: 10.1109/TGRS.2005.856106.
- [62] J. Liu and S. Liang, “Pan-sharpening using a guided filter,” *Int. J. Remote Sens.*, vol. 37, no. 8, pp. 1777–1800, Apr. 2016. doi: 10.1080/01431161.2016.1163749.
- [63] M. Ghahremani and H. Ghassemian, “A new merging method and its spectral and spatial effects,” *Int. J. Remote Sens.*, vol. 36, no. 16, pp. 4131–4143, 2015. doi: 10.1080/01431161.2015.1071897.
- [64] W. Liao et al., “Two-stage fusion of thermal hyperspectral and visible RGB image by PCA and guided filter,” in *Proc. 2015 7th Workshop Hyperspectral Image Signal Processing: Evolution Remote Sensing (WHISPERS)*, pp. 1–4. doi: 10.1109/WHISPERS.2015.8075405.
- [65] J. Duran, A. Buades, B. Coll, C. Sbert, and G. Blanchet, “A survey of pansharpening methods with a new band-decoupled variational model,” *ISPRS J. Photogramm. Remote Sens.*, vol. 125, pp. 78–105, Mar. 2017. doi: 10.1016/j.isprsjprs.2016.12.013.
- [66] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, “A variational model for P+XS image fusion,” *Int. J. Comput. Vision*, vol. 69, no. 1, pp. 43–58, Aug. 2006. doi: 10.1007/s11263-006-6852-x.
- [67] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, “A new pansharpening algorithm based on total variation,” *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 318–322, Jan. 2014. doi: 10.1109/LGRS.2013.2257669.
- [68] X. He, L. Condat, J. Bioucas-Dias, J. Chanussot, and J. Xia, “A new pansharpening method based on spatial and spectral sparsity priors,” *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4160–4174, Sept. 2014. doi: 10.1109/TIP.2014.2333661.
- [69] H. A. Aly and G. Sharma, “A regularized model-based optimization framework for pan-sharpening,” *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2596–2608, Apr. 2014. doi: 10.1109/TIP.2014.2316641.
- [70] M. Möller, T. Wittman, A. L. Bertozzi, and M. Burger, “A variational approach for sharpening high dimensional images,” *SIAM J. Imag. Sci.*, vol. 5, no. 1, pp. 150–178, 2012. doi: 10.1137/100810356.
- [71] G. Zhang, F. Fang, A. Zhou, and F. Li, “Pan-sharpening of multi-spectral images using a new variational model,” *Int. J. Remote Sens.*, vol. 36, no. 5, pp. 1484–1508, 2015. doi: 10.1080/01431161.2015.1014973.
- [72] F. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, “Model-based reduced-rank pansharpening,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 656–660, Apr. 2020. doi: 10.1109/LGRS.2019.2926681.
- [73] P. Liu, L. Xiao, and T. Li, “A variational pan-sharpening method based on spatial fractional-order geometry and spectral-spa-

- tial low-rank priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, Mar. 2018. doi: 10.1109/TGRS.2017.2768386.
- [74] L. Deng, G. Vivone, W. Guo, M. D. Mura, and J. Chanussot, "A variational pansharpening approach based on reproducible kernel Hilbert space and heaviside function," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4330–4344, Sept. 2018. doi: 10.1109/TIP.2018.2839531.
- [75] C.-H. Wang, C.-H. Lin, J. Bioucas-Dias, W.-C. Zheng, and K.-H. Tseng, "Panchromatic sharpening of multispectral satellite imagery via an explicitly defined convex self-similarity regularization," in *Proc. 2019 IEEE Int. Geoscience Remote Sensing Symp. (IGARSS 2019)*, pp. 3129–3132. doi: 10.1109/IGARSS.2019.8900610.
- [76] T. Wang, F. Fang, F. Li, and G. Zhang, "High-quality Bayesian pansharpening," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 227–239, Aug. 2019. doi: 10.1109/TIP.2018.2866954.
- [77] D. Fasbender, J. Radoux, and P. Bogaert, "Bayesian data fusion for adaptable image pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1847–1857, June 2008. doi: 10.1109/TGRS.2008.917131.
- [78] Y. Zhang, S. D. Backer, and P. Scheunders, "Noise-resistant wavelet-based bayesian fusion of multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3834–3843, Nov. 2009. doi: 10.1109/TGRS.2009.2017737.
- [79] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model-based fusion of multi- and hyperspectral images using PCA and wavelets," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2652–2663, May 2015. doi: 10.1109/TGRS.2014.2363477.
- [80] Y. Zhang, A. Duijster, and P. Scheunders, "A Bayesian restoration approach for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3453–3462, Sept. 2012. doi: 10.1109/TGRS.2012.2184122.
- [81] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011. doi: 10.1109/TGRS.2010.2067219.
- [82] C. Jiang, H. Zhang, H. Shen, and L. Zhang, "A practical compressed sensing-based pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 629–633, July 2015. doi: 10.1109/LGRS.2011.2177063.
- [83] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4779–4789, Sept. 2013. doi: 10.1109/TGRS.2012.2230332.
- [84] M. Cheng, C. Wang, and J. Li, "Sparse representation based pansharpening using trained dictionary," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 293–297, Jan 2014. doi: 10.1109/LGRS.2013.2256875.
- [85] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, May 2013. doi: 10.1109/TGRS.2012.2213604.
- [86] X. X. Zhu, C. Grohnfeld, and R. Bamler, "Exploiting joint sparsity for pansharpening: The J-SparseFI algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2664–2681, May 2016. doi: 10.1109/TGRS.2015.2504261.
- [87] H. Yin, "Pan-guided cross-resolution projection for local adaptive sparse representation based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4938–4950, July 2019. doi: 10.1109/TGRS.2019.2894702.
- [88] M. R. Vicinanza, R. Restaino, G. Vivone, M. Dalla Mura, G. Licciardi, and J. Chanussot, "A pansharpening method based on the sparse representation of injected details," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 180–184, Jan. 2015. doi: 10.1109/LGRS.2014.2331291.
- [89] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang, "A new pansharpening method with deep neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1037–1041, May 2015. doi: 10.1109/LGRS.2014.2376034.
- [90] J. Zhong, B. Yang, G. Huang, F. Zhong, and Z. Chen, "Remote sensing image fusion with convolutional neural network," *Sens. Imag.*, vol. 17, no. 1, pp. 1–16, June 2016. doi: 10.1007/s11220-016-0135-6.
- [91] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, pp. 594:1–594:22, July 2016. doi: 10.3390/rs8070594.
- [92] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017. doi: 10.1109/LGRS.2017.2736020.
- [93] Y. Rao, L. He, and J. Zhu, "A residual convolutional neural network for pan-sharpening," in *Proc. Int. Workshop Remote Sensing with Intelligent Processing (RSIP)*, pp. 1–4, May 2017. doi: 10.1109/RSIP.2017.7958807.
- [94] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. 2017 IEEE Int. Conf. Computer Vision (ICCV)*, pp. 1753–1761. doi: 10.1109/ICCV.2017.193.
- [95] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sept. 2018. doi: 10.1109/TGRS.2018.2817393.
- [96] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multi-scale and multidepth convolutional neural network for remote sensing imagery pan-sharpening," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 11, no. 3, pp. 978–989, Mar. 2018. doi: 10.1109/JSTARS.2018.2794888.
- [97] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 11, no. 5, pp. 1656–1669, May 2018. doi: 10.1109/JSTARS.2018.2805923.
- [98] W. Yao, Z. Zeng, C. Lian, and H. Tang, "Pixel-wise regression using U-Net and its application on pansharpening," *Neurocomputing*, vol. 312, pp. 364–371, Oct. 2018. doi: 10.1016/j.neucom.2018.05.103.
- [99] X. Liu, Y. Wang, and Q. Liu, "PSGAN: A generative adversarial network for remote sensing image pan-sharpening," in *Proc. 2018 25th IEEE Int. Conf. Image Processing (ICIP)*, pp. 873–877. doi: 10.1109/ICIP.2018.8451049.
- [100] Y. Zhang, C. Liu, M. Sun, and Y. Ou, "Pan-sharpening using an efficient bidirectional pyramid network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5549–5563, Aug. 2019. doi: 10.1109/TGRS.2019.2900419.

- [101] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3D full convolutional neural network," *Remote Sens.*, vol. 9, no. 11, pp. 1139:1–1139:22, Nov. 2017. doi: 10.3390/rs9111139.
- [102] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 305–319, Dec. 2018. doi: 10.1016/j.isprsjprs.2018.09.018.
- [103] M. Gargiulo, A. Mazza, R. Gaetano, G. Ruello, and G. Scarpa, "Fast super-resolution of 20 m sentinel-2 bands using convolutional neural networks," *Remote Sens.*, vol. 11, no. 22, pp. 2635:1–2635:18, Nov. 2019. doi: 10.3390/rs11222635.
- [104] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse Dirichlet-Net for hyperspectral image super-resolution," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pp. 2511–2520, 2018.
- [105] Y. Qu, H. Qi, and C. Kwan, "Unsupervised and unregistered hyperspectral image super-resolution with mutual Dirichlet-Net," 2019, arXiv 1904.12175.
- [106] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016. doi: 10.1109/TPAMI.2015.2439281.
- [107] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conf. Computer Vision Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.
- [108] S. Vitale and G. Scarpa, "A detail-preserving cross-scale learning strategy for CNN-based pansharpening," *Remote Sens.*, vol. 12, no. 3, pp. 348:1–348:20, Jan. 2020. doi: 10.3390/rs12030348.
- [109] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Inform. Fusion*, vol. 62, pp. 110–120, Oct. 2020. doi: 10.1016/j.inffus.2020.04.006.
- [110] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, June 1997.
- [111] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti, and L. Wald, "Image fusion - the ARSIS concept and some successful implementation schemes," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, nos. 1–2, pp. 4–18, June 2003. doi: 10.1016/S0924-2716(03)00013-3.
- [112] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008. doi: 10.1109/TGRS.2007.912448.
- [113] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Quantitative quality evaluation of pansharpened imagery: Consistency versus synthesis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1247–1259, Mar. 2016. doi: 10.1109/TGRS.2015.2476513.
- [114] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, May 1998. doi: 10.1080/014311698215973.
- [115] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008. doi: 10.14358/PERS.74.2.193.
- [116] M. M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3880–3891, Nov. 2009. doi: 10.1109/TGRS.2009.2029094.
- [117] B. Aiazzi, L. Alparone, S. Baronti, R. Carlà, A. Garzelli, and L. Santurri, "Full-scale assessment of pansharpening methods and data products," in *Image and Signal Processing for Remote Sensing XX*, vol. 9244, L. Bruzzone, Ed. Bellingham, WA: SPIE, 2014, pp. 1–12.
- [118] C. Kwan, B. Budavari, A. C. Bovik, and G. Marchisio, "Blind quality assessment of fused WorldView-3 images by using the combinations of pansharpening and hypersharpening paradigms," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1835–1839, Oct. 2017. doi: 10.1109/LGRS.2017.2737820.
- [119] O. A. Agudelo-Medina, H. D. Benitez-Restrepo, G. Vivone, and A. Bovik, "Perceptual quality assessment of pan-sharpened images," *Remote Sens.*, vol. 11, no. 7, p. 877, Apr. 2019. doi: 10.3390/rs11070877.
- [120] R. Carla, L. Santurri, B. Aiazzi, and S. Baronti, "Full-scale assessment of pansharpening through polynomial fitting of multiscale measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6344–6355, Dec. 2015. doi: 10.1109/TGRS.2015.2436699.
- [121] G. Vivone, R. Restaino, and J. Chanussot, "A Bayesian procedure for full-resolution quality assessment of pansharpened products," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4820–4834, Aug. 2018. doi: 10.1109/TGRS.2018.2839564.
- [122] G. Vivone, P. Addesso, and J. Chanussot, "A combiner-based full resolution quality assessment index for pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 437–441, Mar. 2019. doi: 10.1109/LGRS.2018.2876629.
- [123] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inform. Fusion*, vol. 2, no. 3, pp. 177–186, Sept. 2001. doi: 10.1016/S1566-2535(01)00036-7.
- [124] T.-M. Tu, P. S. Huang, C.-L. Hung, and C.-P. Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 309–312, Oct. 2004. doi: 10.1109/LGRS.2004.834804.
- [125] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images-II. Channel ratio and 'Chromaticity' transform techniques," *Remote Sens. Environ.*, vol. 22, no. 3, pp. 343–365, Aug. 1987. doi: 10.1016/0034-4257(87)90088-5.
- [126] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, 3rd ed. Orlando, FL: Elsevier, 2007.
- [127] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Advantages of Laplacian pyramids over "à trous" wavelet transforms," in *Proc. SPIE Image Signal Process. Remote Sens. XVIII*, vol. 8537, pp. 853704–1–853704–10, 2012.
- [128] B. Aiazzi, L. Alparone, S. Baronti, R. Carlà, A. Garzelli, and L. Santurri, "Sensitivity of pansharpening methods to

- temporal and instrumental changes between multispectral and panchromatic data sets," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 308–319, Jan. 2017. doi: 10.1109/TGRS.2016.2606324.
- [129] M. González-Audiciana, X. Otazu, O. Fors, and A. Seco, "Comparison between Mallat's and the 'à trous' discrete wavelet transform based algorithms for the fusion of multispectral and panchromatic images," *Int. J. Remote Sens.*, vol. 26, no. 3, pp. 595–614, Feb. 2005. doi: 10.1080/01431160512331314056.
- [130] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice Hall, 1995.
- [131] G. Vivone, R. Restaino, and J. Chanussot, "A regression-based high-pass modulation pansharpening approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 984–996, Feb. 2018. doi: 10.1109/TGRS.2017.2757508.
- [132] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007. doi: 10.1109/TGRS.2007.904923.
- [133] G. Vivone, R. Restaino, and J. Chanussot, "Full scale regression-based injection coefficients for panchromatic sharpening," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3418–3431, July 2018. doi: 10.1109/TIP.2018.2819501.
- [134] M. O. Ulfarsson, F. Palsson, M. Dalla Mura, and J. R. Sveinsson, "Sentinel-2 sharpening using a reduced-rank method," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6408–6420, Sept. 2019. doi: 10.1109/TGRS.2019.2906048.
- [135] P. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton Univ. Press, 2008.
- [136] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a matrix toolbox for optimization on manifolds," *J. Mach. Learn. Res.*, vol. 15, no. 42, pp. 1455–1459, Jan. 2014.
- [137] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 1646–1654. doi: 10.1109/CVPR.2016.182.
- [138] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.
- [139] L. Wald, *Data Fusion: Definitions and Architectures—Fusion of Images of Different Spatial Resolutions*. Paris, France: Les Presses de l'École des Mines, 2002.
- [140] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002. doi: 10.1109/97.995823.
- [141] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 313–317, Oct. 2004. doi: 10.1109/LGRS.2004.836784.
- [142] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 6, no. 4, pp. 662–665, Oct. 2009. doi: 10.1109/LGRS.2009.2022650.
- [143] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 930–934, May 2014. doi: 10.1109/LGRS.2013.2281996.